

THE PROBLEM WITH INTEGER PROGRAMMING

H.P.Williams *

London School of Economics

h.p.williams@lse.ac.uk

* This survey paper was given at a workshop, held at the London School of Economics on June 30 2009, to honour Professor Williams' career and retirement.

First published in Great Britain in 2010
by the Operational Research Group, Department of Management
London School of Economics and Political Science

Copyright © The London School of Economics and Political Science, 2010

The contributors have asserted their moral rights.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, without the prior permission in writing of the publisher, nor be circulated in any form of binding or cover other than that in which it is published.

Produced by:

The London School of Economics and Political Science
Houghton Street
London WC2A 2AE

THE PROBLEM WITH INTEGER PROGRAMMING

H.P. Williams *

London School of Economics

h.p.williams@lse.ac.uk

Abstract

Integer Programming (IP), also known as Discrete Optimisation, is a way of modelling a very wide range of problems involving indivisibilities (eg. Yes/No investment decisions) and non-convexities (eg. economies of scale and fixed cost allocation). Such problems arise in many areas which will be mentioned. However IP demands ingenuity in both building models and solving them. A lot is still not properly understood.

This paper investigates the question. ‘Is IP like Linear Programming (LP)?’ The mathematical and economic properties of IP will be contrasted with LP. It will be suggested that the mathematics and economics of IP are still not properly understood. Many of the results which apply to LP do not apply to IP. It will be asserted that this lack of understanding reveals inadequacies in both the mathematics and economics.

It will be shown that in many (but not all) situations the rounding of an LP solution does not produce satisfactory (feasible or optimal) solutions to an IP. A topical example will be given of political apportionment leading to the Alabama Paradox.

IP is essentially concerned with the intersection of two structures:

- (i) Linear inequalities giving rise to polytopes.
- (ii) Lattices of integer points.

Mathematical and computational methods and results exist for both these structures on their own. However mixing them is like mixing oil and water. Problems arise in both the computation of optimal solutions and the economic interpretation of the results. It will be suggested that the appropriate mathematical structure is an integer monoid. This structure will be explained. Connected with this structure are Chvátal functions which are made up of non-negative combinations of the arguments together with the (nested) integer round-up operation. The use of these functions, in place of the conventional non-negative linear combinations of LP allows one to capture many of the classical LP, results eg. The Weyl-Minkowski theorems, ‘pricing’ of indivisible resources and the closing of the ‘duality gap’ leading to the optimal IP solution. It will be shown how optimal Chvátal functions can be interpreted as the (non-marginal) valuation of indivisible resources. However a major problem remains as to how to represent them in a transparent and compact way.

* This survey paper was given at a workshop, held at the London School of Economics on June 30 2009, to honour Professor Williams’ career and retirement.

1. INTRODUCTION

The general (mixed) Integer Programme (IP) can be written:

$$\text{Maximise/Minimise } \sum_j c_j x_j + \sum_k d_k y_k$$

$$\text{subject to } \sum_j a_{ij} x_j + \sum_k e_{ik} y_k \leq b_i \text{ for all } i$$

$$x_j \geq 0 \text{ for all } j, y_k \geq 0 \text{ and integer for all } k$$

Frequently (but not always) the integer variables are restricted to values 0 and 1 representing (indivisible) Yes/No decisions. In such cases we can view such models as *Logical* statements about a series of Linear Programmes (LPs). 0-1 IPs can be interpreted as *Disjunctions* of LPs. The relationship between Logic and IP is explored by Hooker[15] and by Williams[33].

However we give little attention to this aspect of IP in this paper. Instead we concentrate on the general case.

There are a vast number of actual and potential applications of IP. A survey of many practical applications is given in Appa, Pitsoulis and Williams[1]. Williams[32] concentrates on the modelling of many practical applications. We summarise the major applications below.

1. Extensions of LPs, also to model investment and scheduling decisions in eg Manufacturing, Distribution, Petroleum, Gas and Chemicals give rise to mixed IPs.
2. Non-convex, non-linear optimisation problems can only be satisfactorily solved to (global) optimality using IP methods.
3. Power system loading problems which take account of the fixed costs of switching on generators.
4. Distribution problems which involve routing decisions as well as the location of manufacturing and storage plants.
5. Telecommunications involving the location of eg concentrators as well as bandwidth allocation.
6. Medical radiation problems involving the concentration and direction of x-rays.
7. Statistical design involving the creation of eg Latin Squares.
8. Molecular biology such as eg genome sequencing where fragments of DNA have to be ordered into the most likely sequence.
9. Archaeological seriation where the most likely sequence of sites over time is sought.

The latter two applications are examples of the famous Travelling Salesman Problem (see [18] although these two applications are not discussed there).

10. Minimal size logical statements which can be used in eg medical diagnosis and credit scoring.
11. Computer design where it is desired to locate components with possible objectives of eg minimising wire length (to optimise speed), cross-overs or number of components.
12. Aircraft scheduling ie deciding which aircraft fly on which routes.

13. Aircrew scheduling ie deciding which crews fly on which routes to eg minimise total number of crews or total cost. Constraints take account of physical incompatibilities (ie limitations of time and place) as well as total work time and rest breaks.

It should be emphasised that, while major advances have been made in both methods of modelling and computer systems (see eg Bixby et al [3]) the solving of large examples of some applications remain more aspirations than standard realisations. This is in contrast to LP where models with hundreds of millions of variables and constraints (these arise in LP formulations of Stochastic Programmes, see eg [13]) can be solved.

The reasons for this contrast between LP and IP is one of the major themes of this paper.

2. IP AND ROUNDING

A common reaction of non-specialists when thinking about IP is to consider rounding the optimal LP solution. There are a few examples where rounding can be used (eg the Satisfiability Problem with Horn Clauses in Logic, see eg Williams[33]). The following small example demonstrates why this may be unhelpful.

Consider the LP

Minimise	x_2	with optimal solution
	$2x_1 + x_2 \geq 13$	
subject to	$5x_1 + 2x_2 \leq 30$	$x_1 = 2 \frac{2}{3}, x_2 = 7 \frac{2}{3}$
	$-x_1 + x_2 \geq 5$	
	$x_1, x_2 \geq 0$	

The corresponding IP is

Minimise	x_2	with optimal solution
	$2x_1 + x_2 \geq 13$	
subject to	$5x_1 + 2x_2 \leq 30$	$x_1 = 2, x_2 = 9$
	$-x_1 + x_2 \geq 5$	
	$x_1, x_2 \geq 0$	
	and integer	

There is no obvious way of generally moving from the LP solution to the IP solution.

A number of other contrasts between the LP and IP solutions are worth noting.

- (i) The LP has **3** variables positive (x_1, x_2 and the slack in the 2nd constraint) in the optimal solution. This is a result of the LP having **3** constraints (apart from the non-negativities).

(ii) The LP has **2** constraints *binding* (constraints 1 and 3) in the sense that if either or both of these constraints is removed the optimal solution alters. If constraint 2 is removed the solution is unaltered. Hence constraint 2 has no economic value. This is a result of the LP having **2** structural variables (the stated variables-not the slack or surplus variables in the constraints). The optimal LP solution is at X, the intersection of constraints 1 and 3. If either of these constraints is removed the optimal solution moves to B or C respectively.

These results are easily seen in figure 1.

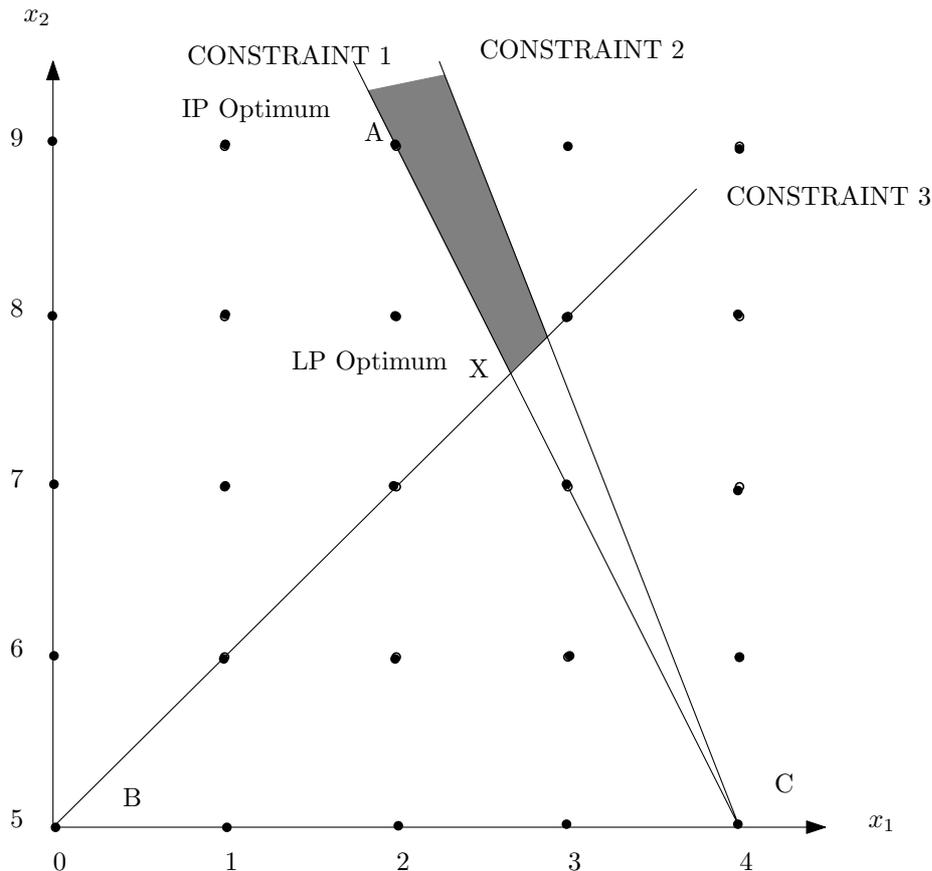


Figure1. LP and IP Optima

But

(iii) The IP optimal solution at A ($x_1 = 2, x_2 = 9$) has 4 variables positive (x_1, x_2 and the slacks in the 2nd and 3rd constraints). There is no simple relationship between this and the size of the IP.

(iv) The IP has 3 constraints *binding* (constraints 1, 2 and 3). Again there is no simple relationship with the size of the IP.

These results are seen in figures 2, 3 and 4.

If constraint 1 is removed the optimal IP solution moves to B.

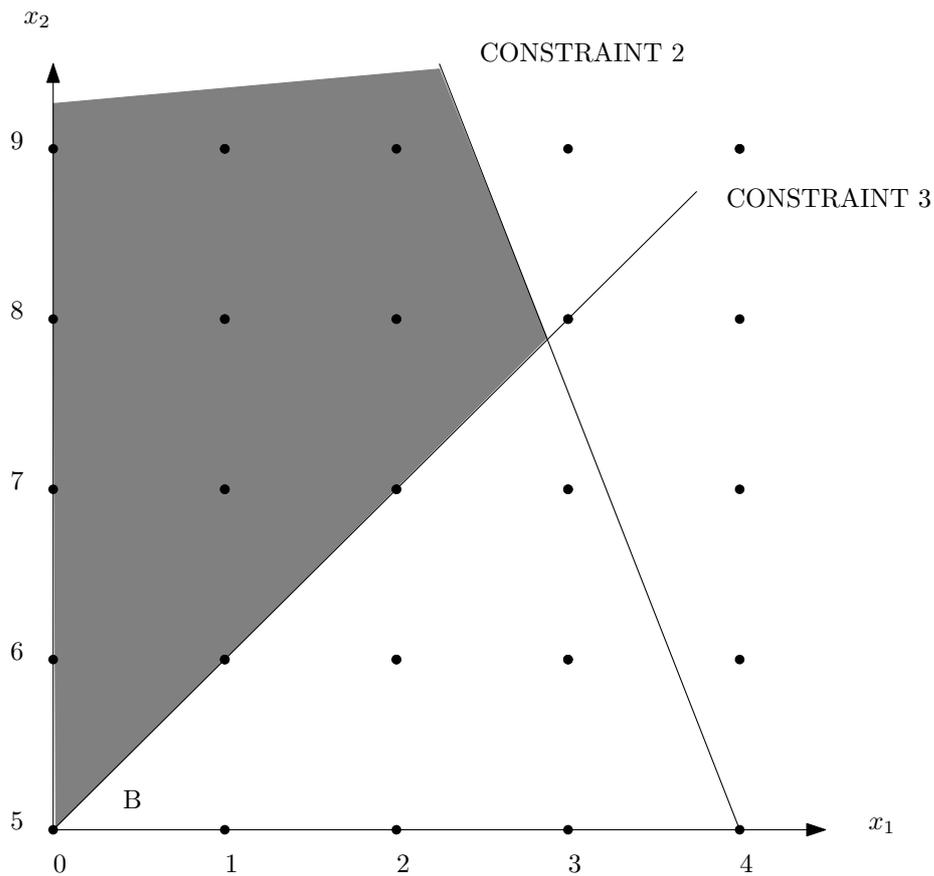


Figure 2. Optimal IP Solution when Constraint 1 Removed

If constraint 2 is removed the optimal IP solution moves to D at $(x_1=8, x_2=3)$. (Constraint 2 is **not** redundant).

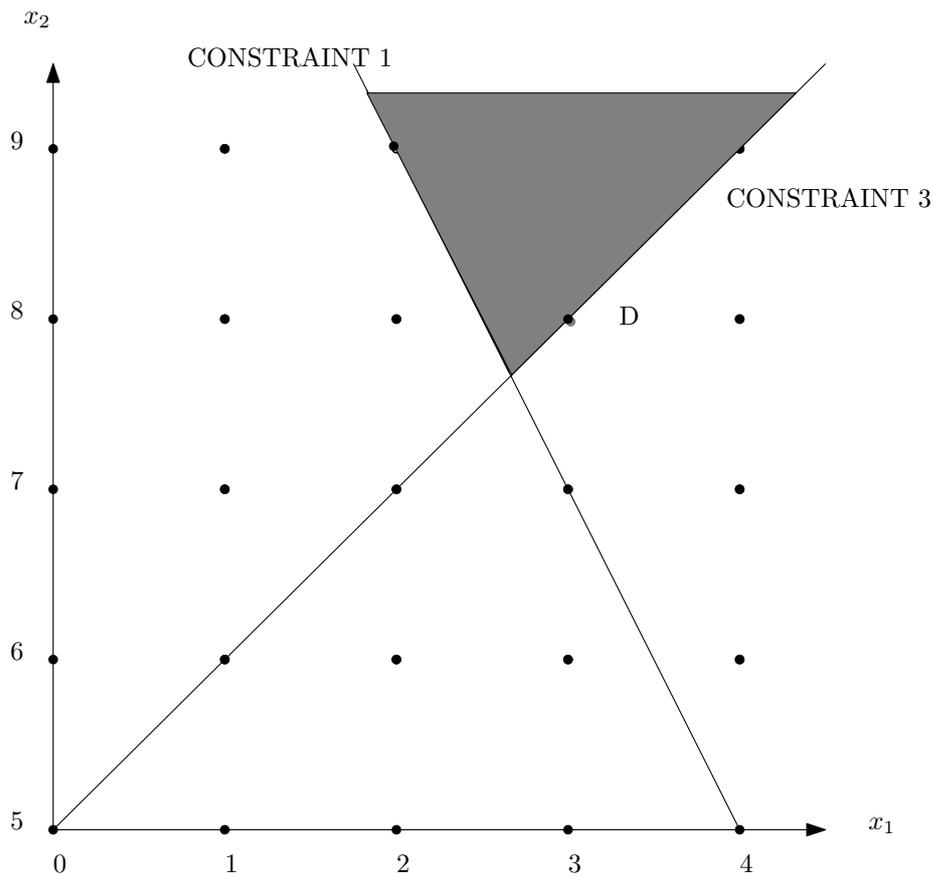


Figure3. Optimal IP Solution when Constraint 2 Removed

If constraint 3 is removed the optimal IP solution moves to C.

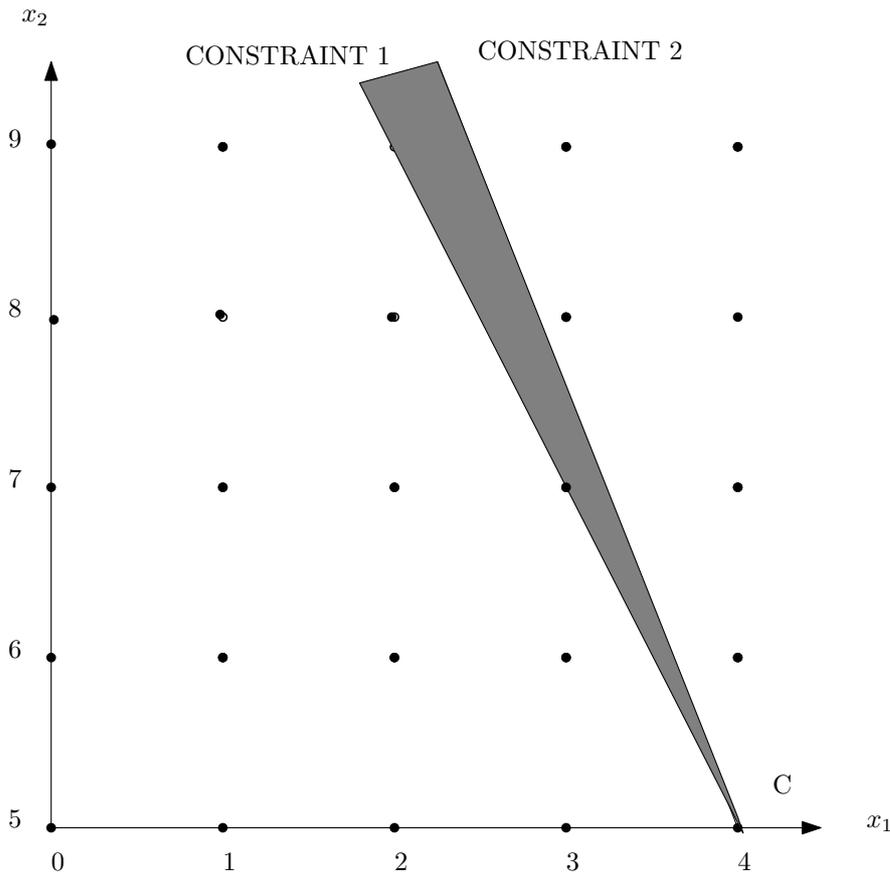


Figure 4. Optimal IP Solution when Constraint 3 Removed

These relationships between LP and IP (or lack of them) are examined, in general, in section 3.

Since the altering of any of the constraints in this IP example (although not all satisfied as equalities) alters the optimal objective value they all have *economic value*. This cannot be measured by *marginal values* as is done by LP dual values. However we seek an analogous IP valuation. This is done in section 4.

Before considering these aspects we further demonstrate the inadequacies of rounding by a practical example.

2.1 The Alabama Paradox

This is concerned with political apportionment. The problem arose in the USA in the 18th century when the number of members of the House of Representatives was increased from 299 to 300, as a result of the method used to allocate representatives to individual states, based on their populations. The number of representatives allocated to Alabama dropped from 8 to 7 (although there had been no changes in populations). A small example demonstrates the method used and why it can lead to such a paradoxical result.

Suppose we have three states with the populations given and 10 representatives to be allocated.

State	Population	Fractional Allocation	Rounded Allocation
A	621k	4.41	4
B	587k	4.17	4
C	201k	1.43	2

The fractional allocations are the proportions of the populations multiplied by 10. The rounded (integer) allocations are obtained by allocating the integer part and then allocating the total shortfall of representatives in decreasing order of fractional component. In this case, after allocating 4, 4 and 1 representatives respectively the remaining representative is allocated to C, having the largest fractional component.

If, however, we increase the total number of representatives to 11 the following allocation results.

State	Population	Fractional Allocation	Rounded Allocation
A	621k	4.81	5
B	587k	4.58	5
C	201k	1.57	1

After allocating 9 of the 11 representatives the remaining 2 are allocated to A and B respectively, having the largest fractional components.

The net effect is that although the total number of representatives has increased C's representation has decreased. This paradox and different methods of apportionment are discussed by Balinski and Young [2].

The reason this example is topical is that it could equally well be applied to allocating seats to *political parties* in a proportional representation system. In the 2009 elections, for the European Parliament, Britain was divided into 12 regions (Northern Ireland used a different system) and electors, in each region, voted for a party, not an individual. Representatives were allocated to parties, depending on how many votes they got. If it were done in the same manner as that above the Alabama paradox could result (the method is said to be 'non-monotonic'). More seriously the 'fairness' of the method could be questioned. The method actually used is the *D'Hondt* system (named after the Belgian mathematician D'Hondt, see [9]).

The D'Hondt system proceeds as follows:

Allocate the first representative to the state (or party) with the most votes (or population). Then, notionally, divide this state's (party's) votes (or population) by 2 and repeat the procedure. When a state (or party) has been allocated r representatives divide the original votes (or population) by $r+1$. Continue in this manner until all representatives have been allocated. After the Alabama paradox was revealed, the USA system was altered to the *Jefferson* method which turns out to be equivalent to the D'Hondt method. However the rationale behind both systems is, at first sight, obscure.

The reason for using this example in this paper is to formulate it as an IP. One of the merits of building an LP or an IP model, to solve a problem, is that it forces one to specify *objectives*. In the case of political apportionment one would probably specify an objective of *fairness*. If one defines fairness in a Rawlsian sense (see Rawls [20]) then one can seek to *Minimise the Maximum* ratio of number representatives to votes (or population).

In order to do this we define *integer* variables x_i which give the number of representatives which should be allocated to state (or party) i .

If state (party) i has population (vote) v_i then we wish to choose x_i so as to

Minimise Maximum $_i (x_i/v_i)$

subject to $\sum_i x_i = \text{Total number of representatives}$

ie make the maximum disparity in ratio of representatives to population (votes) as small as possible, in order to seek to equalise them. (It makes no sense to maximise the minimum ratio, in the political case, since this is likely to be 0 if 'eccentric' parties compete).

The above model can easily be converted into a conventional (mixed) IP

Minimise y

subject to $(x_i/v_i) \leq y$ for all i

$\sum_i x_i = \text{Total number of representatives}$

where y is a continuous variable.

If this model is applied to the above example the following solution results, if there are 10 representatives.

State	Population	Fractional Allocation	IP Model Allocation
A	621k	4.41	5
B	587k	4.17	4
C	201k	1.43	1

If one solves this model as an LP (dropping the integrality conditions to give the ‘LP Relaxation’) one obtains the fractional allocation (where all ratios are equal).

To see that the solution above is ‘better’ (‘fairer’) than the rounded solution (ie rounding the LP solution) one can observe that the maximum ratio, in the solution above, is $5/621$ which is smaller than the maximum ratio of $2/201$ in the rounded solution. Indeed $5/621$ is the minimum possible.

The above solution is that which would be obtained by the D’Hondt (Jefferson) method. It can be shown that this will always be the case. Hence these methods can be justified by the use of the above (‘fair’) IP formulation. In fact (although not usually explained as such) the D’Hondt method is an iterative method of obtaining the Minimax solution).

It is interesting to note what the result would have been if the 2009 European elections (see [10]) had used rounding rather than the fairer D’Hondt (IP) system. The results (in representatives) would have been very different. It would have resulted in (over Britain as a whole)

Party	D’Hondt Solution	Rounded Solution
Conservative	25	21
United Kingdom Independence	13	9
Labour	13	13
Liberal Democrats	11	10
British National	2	7
Green	2	6
Scottish Nationalist	2	2
Plaid Cymru	1	1

3. MATHEMATICAL DIFFERENCES BETWEEN IP AND LP

We list below some of the main differences between the *structure* of the solutions to LPs and IPs. The differences in Computational Complexity between the two classes of models has been intensely studied (see, for example Nemhauser and Wolsey [19]) The results below also suggest why IPs are usually much more difficult to solve than corresponding sized LPs. A standard reference for the theory of LP is Dantzig [7].

Some of these results were illustrated by the 2-variable example in the previous section.

1. If an LP (in standard form), with m constraints (apart from the non-negativities), has an optimal solution then among the optimal solutions will be one **with at most m variables (structurals and logicals) positive**.

Such a solution is usually known as a *basic* solution, since the columns of the coefficient matrix form a vector space of m -tuples. Such a vector space can have a dimension of at most m . Basic solutions correspond to vertices of the associated polytope (see figure 1).

This result can be exploited in methods such as the Simplex algorithm which restrict their search to the (finite number of) basic solutions.

For IP there is no such corresponding result. Every variable could be positive, irrespective of the size of m . This is demonstrated by the following class of IPs with 1 constraint (known as Knapsack problems).

Take n prime numbers p_1, p_2, \dots, p_n . Define $P = p_1 p_2 \dots p_n$ ie their product. Define $P_i = P/p_i$.

Consider the IP

$$\text{Maximise } \sum_i P_i x_i$$

$$\text{subject to } \sum_i P_i x_i \leq \sum_i P_i$$

$$x_i \geq 0 \text{ and integer for all } i$$

A non-negative combination of any proper subset of the P_i must be a multiple of at least one of the p_i . But the right-hand-side cannot be a multiple of any of the p_i since dividing it by any p_i leaves a positive remainder. Hence the constraint cannot be satisfied as equality by any other than the solution $x_i = 1$ for all i . This is therefore the unique optimal solution showing all variables must be positive.

2. **For an LP, with n structural variables, at most n constraints (including non-negativity constraints) will be binding** (in the sense that removing them would cause the optimal objective value to change).

This result is equivalent to 1 above, for LP, since positive dual values correspond to binding constraints.

For an IP, with n structural variables, at most $2^n - 1$ constraints (non-negativity constraints) will be binding. This result is due to Scarf [24].

3. There are valuations on the constraints of an LP model which close the duality gap ie there is a (symmetric) LP (dual) model. The binding constraints have positive dual valuations. These dual values (when unique) represent the effect, on the optimal objective value, of marginal changes in the right-hand-sides of the constraints and therefore have considerable economic importance in many applications.

For an IP model there are no corresponding valuations which close the duality gap. However there are *Chvátal functions* (see Chvátal [6] , discussed in section 5) which close the duality gap.

There is no obviously symmetric dual model which gives rise to the optimal Chvátal function.

The results above suggest the, rather unsatisfactory, understanding of the structure of solutions of (pure) IPs compared with LPs. They suggest a lack of understanding of the mathematical nature of IP. Misconceptions exist regarding structural properties of IP (and LP) such as those above. These are well covered by Greenberg [14].

4. Lattices within Polytopes

One of the problems with understanding the mathematical nature of IP is that there are essentially two structures involved. The conventional LP constraints give rise to **polytopes**, but within these polytopes the integrality conditions give rise to **lattices** (this is generally the integer lattice, but reformulations can result in other lattices). For LP the, well known, structural results discussed above exist. Analogous results exist for lattices. When, however, the two structures are combined many of these structural results are lost.

We consider the model below.

$$\begin{array}{ll}
 \text{Maximise} & 2x_1 + x_2 \\
 & 2x_1 + 9x_2 \leq 80 \\
 & 2x_1 - 3x_2 \leq 6 \\
 & -x_1 \leq 0 \\
 \text{subject to} & -x_2 \leq 0 \\
 & 2x_1 + 3x_2 \equiv 0 \pmod{12} \\
 & x_1 \equiv 0 \pmod{1} \\
 & x_2 \equiv 0 \pmod{1}
 \end{array}$$

The first four constraints define a polytope and the latter three constraints a lattice.

In order to illustrate a duality result for both structures independently we will firstly ask what is the strongest statement about the objective function implied by the first four constraints. This may be obtained from the optimal dual values of $\frac{1}{3}, \frac{2}{3}, 0, 0$ on the constraints. Adding them together in these multiples gives $2x_1 + x_2 \leq 30\frac{2}{3}$, which is the optimal LP objective value.

The strongest statement about the objective function implied by the latter three constraints can also be obtained from multipliers on the constraints. Taking optimal multipliers of 3, 0, 4 on the constraints and adding them together gives $2x_1 + 13x_2 \equiv 0 \pmod{4}$. Since $13 \equiv 1 \pmod{4}$ we have $2x_1 + x_2 \equiv 0 \pmod{4}$. 4 is the largest such modulus implied. This duality result for congruences, presented in an analogous fashion to that for LP, is given by Williams [26] and [28]. It is sometimes known as

the integer form of Farkas' Lemma (see Schrijver [25]) and is equivalent to a result of Kronecker [17].

Taken together the two deductions, by means of multipliers, can only allow us to deduce that $2x_1 + x_2 \leq 28$. But figure 5 demonstrates that the optimal objective value, subject to the two types of constraints, is 24 since the line $2x_1 + x_2 = 28$ intersects no lattice points **within** the polytope.

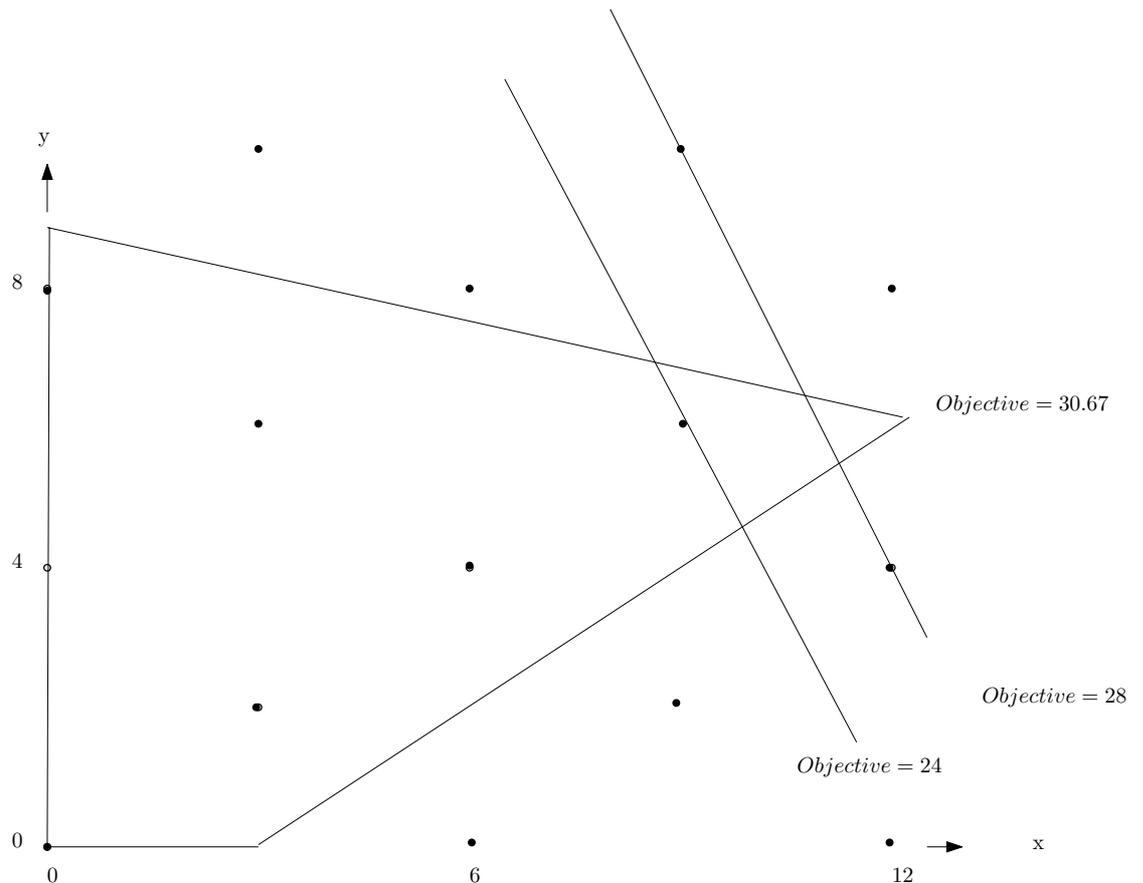


Figure 5. A Lattice within a Polytope

Technically all we can deduce from the duality result for the congruences, taken together with the duality result for the LP Relaxation, is a *rank 1 cut* (see eg Nemhauser and Wolsey [19]). The combining of the two structures demands more sophisticated analysis (for which there are few structural results).

It is worth noting that the solution of the inequality system can be done by the Simplex algorithm (see eg Dantzig [7]) and the 'solution' of the congruence system can be done by the Euclidean algorithm (see Williams [26]). These are possibly the two most powerful algorithms in mathematics.

A polytope can be represented as the sum of a polyhedron (a closed polytope) and a cone. A lattice within a polyhedron is obviously represented by a finite number of points. However lattices within

cones give rise to an infinite number of points. They do, however, still have a structure. This structure is an **integer monoid**. We contend that this structure is fundamental to understanding IP.

An integer monoid consists of the non-negative integer combinations of a finite set of generating vectors (its Hilbert basis). Some integer monoids are *polyhedral* and some are non-polyhedral. Polyhedral integer monoids are those which are represented by lattice points within a cone. This is illustrated in figure 6. We consider the cone defined by the inequalities

$$-2x + 7y \geq 0$$

$$x - 3y \geq 0$$

and the lattice points with integer coordinates within this cone.

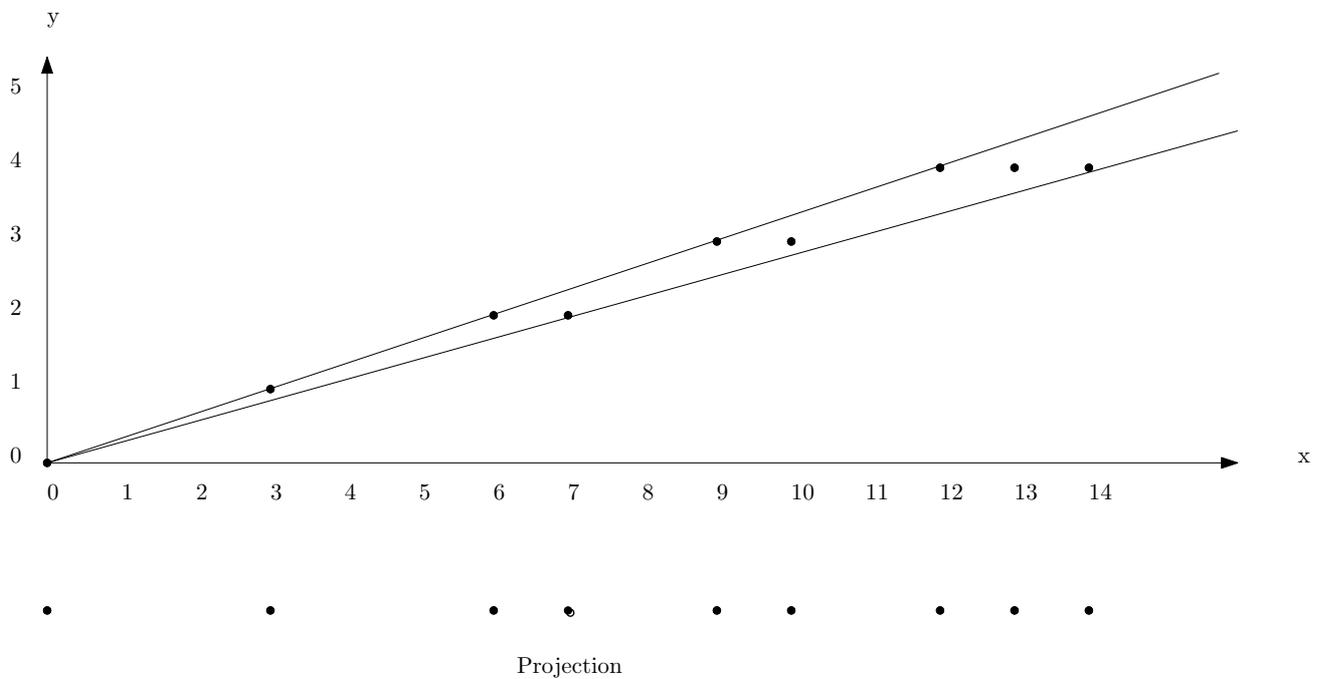


Figure 6. A Polyhedral Monoid and its Projection into a Non-Polyhedral Monoid

The lattice within the cone is the polyhedral monoid generated by the vectors $\begin{pmatrix} 3 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 7 \\ 2 \end{pmatrix}$.

An example of a non-polyhedral monoid is given by the projection of the above monoid onto the x axis. This monoid is generated by the integers 3 and 7. Clearly this monoid is only a subset of the integer points in the cone $x \geq 0$.

Projection converts LPs into LPs, in a lower dimension, and offers a method of solving LPs (see Williams [31]) as well as deriving the dual. For IP this is not the case. We need to optimise over a monoid. However it is always possible to lift a non-polyhedral integer monoid into a polyhedral one in a higher dimension (see Kirby and Williams [16]).

Integer monoids have a structure, although it is not obvious how to exploit this in solving and understanding the structure of IPs. For example 1-dimensional integer monoids, where the generators have a greatest common divisor of 1, eventually generate all integers. The last integer which cannot be generated is known as the Frobenius number (see Rödseth [22]). It can be shown that an integer monoid generated by (coprime) integers a and b has Frobenius number $ab-(a+b)$. The first part of the monoid, up to and including, the Frobenius number, is known as the ‘head’ and the rest, the ‘tail’. It can also be shown that the head is the ‘mirror image’ of itself when reversed, in the sense that the non-representable values mirror the representable ones. This is demonstrated by figure 7, where the projected monoid has been reversed at the bottom with 0 corresponding to the Frobenius number (11).

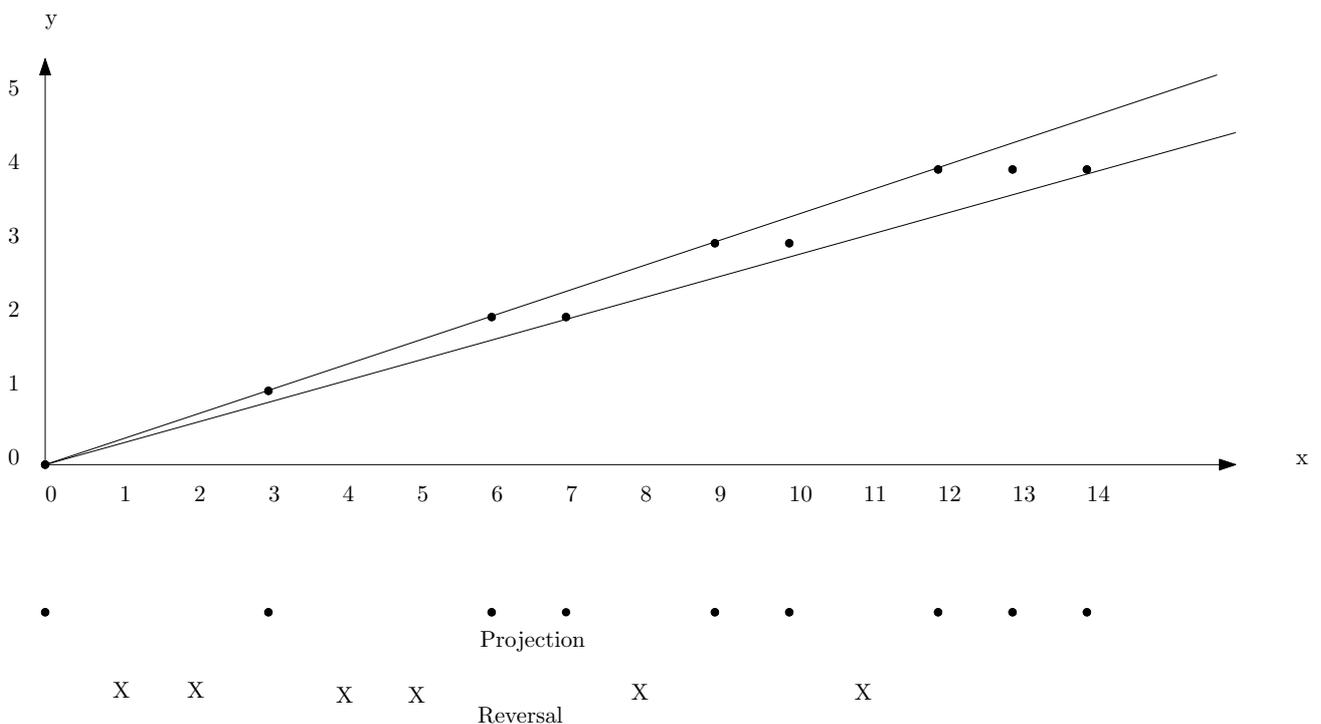


Figure 7. The Head of a Monoid Mirroring its Reversal

Despite their obvious structure, this has not, so far, been exploited in any algorithmic, or interpretative way with regard to IP.

5. Duality in LP and IP

As discussed above LPs have well defined duals which are of great mathematical, structural and economic importance. In fact a number of Nobel prizes for Economics have resulted from the exploitation of this concept in various contexts (eg Leontief, Samuelson, Koopmans, Kantorovitch), (see eg Dorfman, Samuelson and Solow [8]).

Valuations on discrete resources would have a number of applications. This question was first addressed by Gomory and Baumol [12]. In particular an ‘unsolved’ problem of accountancy is how to allocate fixed costs in an allocatively efficient fashion (see eg Butler and Williams [5]). We illustrate this by a model for the facilities location problem.

Let $y=0$ or 1 depending on whether a facility should, or should not, be built, at a fixed cost of f .

x_i is the level of service to be provided to consumer i (up to level D_i) at a profit of p_i per unit provided.

A (mixed) IP model is

$$\begin{aligned} & \text{Maximise} && \sum_i p_i x_i - fy \\ & \text{subject to} && x_i - D_i y \leq 0 \text{ for all } i \end{aligned}$$

A dual value of v_i on $x_i - D_i y \leq 0$ would result in

$$\text{Maximise} \quad \sum_i (p_i - v_i)x_i - (f - \sum_i D_i v_i)y$$

ie an allocation of the fixed cost back to the consumers.

No obvious corresponding IP result is known to match LP duality.. However we construct what we regard as the most satisfactory structure for (pure) IPs by means of the *value function* of an IP. The value function of an LP or an IP is the optimal objective value as a function of the right-hand-side values. For an LP this can be constructed from the duality theorem using knowledge of all the vertices of the dual model. We illustrate this, using the numerical example from section 2, with general right-hand-side values. This is

$$\begin{aligned} & \text{Minimise} && x_2 \\ & \text{subject to} && 2x_1 + x_2 \geq b_1 \\ & && 5x_1 + 2x_2 \leq b_2 \\ & && -x_1 + x_2 \geq b_3 \\ & && x_1, x_2 \geq 0 \end{aligned}$$

The value function is

$$\text{Max } (5b_1 - 2b_2, \frac{1}{3}b_1 + \frac{2}{3}b_3, b_3)$$

The LP is feasible so long as

$$\text{Max } (2b_1 - b_2, -b_2, -b_2 + 2b_3) \leq 0$$

This latter expression is known as the *consistency tester*.

The coefficients in the value function terms $((5, -2, 0), (\frac{1}{3}, 0, \frac{2}{3}), (0, 0, 1))$ represent the vertices of the dual polytope. This is because any right-hand-side for the original model corresponds to a set of objective coefficients in the dual model. Multiplying these by each vertex in the dual model, and taking the maximum, gives, by the LP duality theorem, the common optimal value for both models.

The coefficients in the consistency tester terms $((2, -1, 0), (0, -2, 0), (0, -1, 2))$ represent the extreme rays of the dual polytope. This is because if they are multiplied by any objective coefficients of the dual model to make it unbounded, the corresponding primal model with these coefficients, as right-hand-side, must be infeasible.

We can easily check that this LP is feasible with the values for the example in section 2, $b_1 = 13$, $b_2 = 30$, $b_3 = 5$ and that the optimal objective value is

$$\text{Max } (5, 7\frac{2}{3}, 5) = 7\frac{2}{3}$$

In order to try to capture the reliance of the LP value function on the duality results we consider the corresponding value function for the (pure) IP (see Blair and Jeroslow [4]).

$$\begin{array}{ll} \text{Minimise} & x_2 \\ & 2x_1 + x_2 \geq b_1 \\ \text{subject to} & 5x_1 + 2x_2 \leq b_2 \\ & -x_1 + x_2 \geq b_3 \\ & x_1, x_2 \geq 0 \\ & \text{and integer} \end{array}$$

The value function is

$$\text{Max } (5b_1 - 2b_2, \lceil \frac{1}{3}(b_1 + 2b_3) \rceil, b_3, b_1 + 2 \lceil \frac{1}{5}(-b_2 + 2 \lceil \frac{1}{3}(b_1 + 2b_3) \rceil) \rceil)$$

where $\lceil \cdot \rceil$ represents the integer round-up operation.

The consistency tester is (in this instance) the same as for the LP.

The above expression is known as a *Gomory function* since it reflects Gomory's algorithm for PIPs (see Gomory [11]). The component terms, involving linear combinations together with (in general nested) integer round-up operations, are known as *Chvátal functions* after Chvátal [6].

Using the values $b_1 = 13$, $b_2 = 30$, $b_3 = 5$ the optimal objective value is

$$\text{Max}(5, 8, 5, 9) = 9.$$

Hence the Chvátal function $b_1 + 2 \lceil \frac{1}{5}(-b_2 + 2 \lceil \frac{1}{3}(b_1 + 2b_3) \rceil) \rceil$ is that which gives rise to the optimal objective value. This confirms the observation illustrated in figures 2, 3 and 4 that all three constraints are binding and should therefore be given an 'economic valuation'. This aspect is discussed by Williams [29]. But the Chvátal function above cannot be reduced (as in the LP case) to a linear expression with coefficients corresponding to dual (marginal) values. Nor does the 'relaxation' of this Chvátal function (dropping the round-up operation and collecting terms) correspond to a vertex of the LP polytope.

In general (for a minimisation with \geq constraints) the optimal Chvátal function will consist of a non-negative linear combination of the variables, together with nested integer round up operations. If it is convenient to include $=$ and \leq constraints then these can give rise to negative coefficients at the innermost level (as illustrated by the coefficient of b_2 in the above example).

A Gomory function is a maximisation of a finite set of Chvátal functions. Wolsey [34] refers to this as the 'b-hull of a PIP'.

For maximisations subject to \leq constraints we would redefine Chvátal functions using the integer round-down operation.

Chvátal functions are the IP analogue of the linear functions (using dual values as coefficients) for LP.

Calculating optimal Chvátal functions is computationally difficult (unlike calculating optimal dual values in the LP case). If it turns out that, for an n variable PIP, only those constraints binding at the LP optimum are needed (known as the IP over the optimal LP cone) then the value function reduces to a Chvátal function. The calculation of this Chvátal function is straightforward and is discussed by Williams [27].

Another problem is that of *representing* Chvátal functions in a transparent and unique way ('normal form'). In contrast, for the LP case, all that is needed is the vector of dual values. A useful representation for the PIP case is as a directed *tree*. Each arc and each node has an associated number. Numbers on arcs represent multiples in which associated terms are added and numbers on nodes represent division together with integer round-up. We illustrate the tree associated with the Chvátal function $b_1 + 2 \lceil \frac{1}{5}(-b_2 + 2 \lceil \frac{1}{3}(b_1 + 2b_3) \rceil) \rceil$ in figure 8.

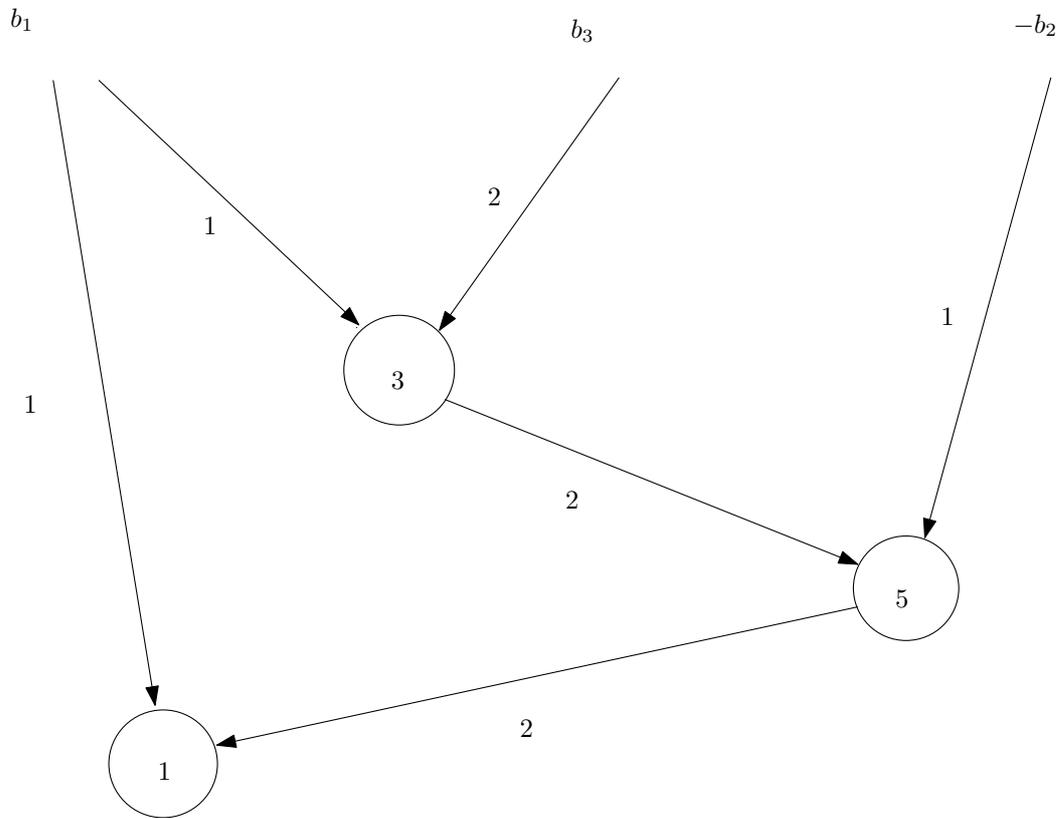


Figure 8. Tree Representation of a Chvátal Function

There is no guarantee that such a representation will be unique. Indeed it is sometimes possible to remove unnecessary rounding operations in Chvátal functions. This is explored, in a limited way, by Rhodes and Williams [21].

For example it is easy to verify, for integer n

$$\lceil \frac{2}{7} \lceil \frac{7}{3} n \rceil \rceil \equiv \lceil \frac{2}{3} n \rceil$$

But

$$\lceil \frac{7}{3} \lceil \frac{2}{7} n \rceil \rceil \neq \lceil \frac{2}{3} n \rceil \quad \text{eg } n = 1$$

$$\lceil \frac{1}{3} \lceil \frac{5}{6}n \rceil \rceil \equiv \lceil \frac{5}{18}n \rceil$$

But

$$\lceil \frac{2}{3} \lceil \frac{5}{6}n \rceil \rceil \neq \lceil \frac{5}{9}n \rceil \quad \text{eg } n = 5$$

Two other properties of Chvátal functions are worth remarking on.

- (i) They obey the *triangle inequality* (when using integer round-up) since

$$\lceil a \rceil + \lceil b \rceil \geq \lceil a + b \rceil$$

Hence they are of use in designing Discrete Metrics (see Rhodes and Williams [21]).

- (ii) They are *Shift Periodic* ie if f is a Chvátal function

$$f(x_1 + a, x_1 + a, x_2 + a, x_2 + a, \dots, x_n + a, x_n + a) = f(x_1, x_2, \dots, x_n) + b$$

for some a_i and b ,

This is illustrated in figure 9 for the Chvátal function of one variable

$$\lceil \frac{1}{2} (x + 3 \lceil \frac{1}{9}x \rceil) \rceil \quad \text{which is } (9, 6) \text{ shift periodic, ie } f(x + 9) = f(x) + 6.$$

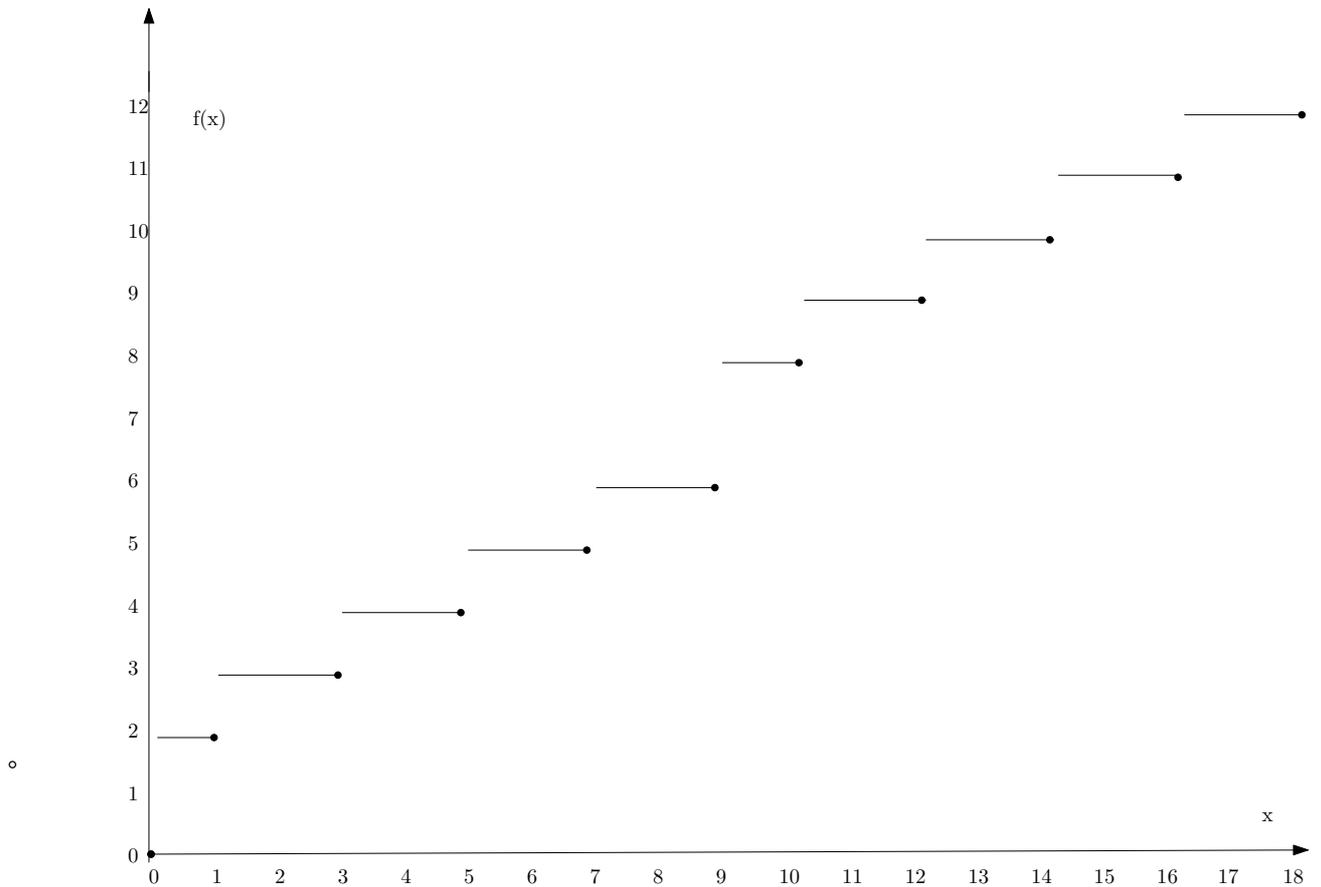


Figure 9. Shift Periodicity of a Chvátal Function

It can be seen that this function is discontinuous, with an average gradient of $2/3$, but ‘repeats itself’ with a period of 9.

Another ‘unification’ which is provided by Chvátal functions concerns the Weyl-Minkowski theorems (see eg Schriver [25]). These give the ‘dual’ results that a polytope can alternatively be described by linear inequalities or as a convex combination of its vertices together with a non-negative combination of its extreme rays.

This extends to polyhedral monoids which can alternatively be described as the integer points within a polytope described by inequalities or as integer multiples of the elements in its Hilbert basis. For example the polyhedral monoid illustrated in figure 6 can be described as the integer points within the inequalities

$$-2x + 7y \geq 0$$

$$x - 3y \geq 0$$

or alternatively as the non-negative integer multiples of the vectors $\begin{pmatrix} 3 \\ 1 \end{pmatrix}$ and

$$\begin{pmatrix} 7 \\ 2 \end{pmatrix}.$$

Although non-polyhedral monoids cannot be described as the integer points satisfying linear inequalities, they can always be described as the integer points satisfying inequalities involving Chvátal functions. The (projected) non-polyhedral monoid, also illustrated in figure 6, can be described as the non-negative integer multiples of 3 and 7. It can alternatively be described as the integer points satisfying

$$\lceil -x/3 \rceil + \lceil 2x/7 \rceil \leq 7$$

This result is given by Ryan and Trotter [23] and Blair and Jeroslow [4].

In conclusion we suggest that integer monoids and Chvátal functions are fundamental structures in integer programming and that the subject should be extended to the problem of optimising the latter over the former. This could give the subject a unity which it so far lacks, by comparison with LP.

6. References

1. Appa, G., Pitsoulis, L. and Williams, H.P. 2006 *Handbook of Modelling for Discrete Optimization*, Springer, New York.
2. Balinski, M.L. and Young, H.P. 1982 *Fair Apportionment*, 2nd Edition, Yale University Press, Yale.
3. Bixby, R.E., E., Fenelon, M., Rothberg, E. and Wunderling, R. 1999 MIP: Theory and practice-closing the gap, *IFIP Conference Proceedings* **174** 19-50 Kluwer, Deventer, Holland.
4. Blair, C.E. and Jeroslow, R. 1982 The value function of an integer program *Mathematical Programming* **23** 237-273
5. Butler, M. and Williams, H.P. 2006 The allocation of shared fixed costs. *European Journal of Operational Research* **170**, 391-397
6. Chvátal, V. 1973 Edmonds polytopes and a hierarchy of combinatorial problems. *Discrete Mathematics* **4** 305-337
7. Dantzig, G.B. 1963 *Linear Programming and Extensions*, Princeton Univ. Press, Princeton, NJ.
8. Dorfman, R., Samuelson, P.A. and Solow, R.M. 1958 *Linear Programming and Economic Analysis*, McGraw Hill, New York.
9. D'Hondt, V. 1882 *Système pratique et raisonne de representation propoortionnelle*, Bruxelles: Librarie C. Muquardt.
10. European Election Results, <http://www.europarl.org.uk/european-elections>.
11. Gomory, R.E. 1963 An algorithm for integer solutions to linear programs. In *Recent Advances in Mathematical Programming*, R. Graves and P. Wolfe (Eds) McGraw Hill 269-302

12. Gomory, R.E. and Baumol, W.J. 1960 Integer programming and pricing *Econometrica* **28** 521-550
13. Gondzio, J. and Grothey, 2005 A. Direct solution of linear systems of size 10^9 arising in optimization and interior point methods. In Wyrzykowski, R., Dongarra, J. Meyer, N. and Wasniewski, J. (Eds) *Parallel Processing and Applied Mathematics PPAM 2005*, Lecture Notes in Computer Science, **3911** Springer-Verlag, New York, 513-525
14. Greenberg, H. 2009 Myths and Counter Examples in Mathematical Programming, web page: <http://glossary.computing.society.informs.org/myths/currentversion/myths.pds>
15. Hooker, J.N. 2000 *Logic-based Methods of Optimization*, Wiley, New York
16. Kirby, D. and Williams, H.P. 1997 Representing integral monoids by inequalities, *Journal of Combinatorial Mathematics and Combinatorial Computing* **23**, 87-95
17. Kronecker, L. 1884 Näherungsweise ganzzahlige Auflösung lineare Gleichungen, *Monatsberichte der Königlich Preussischen Akademie der Wissenschaften zu Berlin* 1179-1193, 1271-1299
18. Lawler, E.L., Lenstra, J.K., Rinnooy, A.H.G., and Shmoys, D.B. (Eds) 1995 *The Travelling Salesman Problem*, Wiley, Chichester.
19. Nemhauser, G.L. and Wolsey, L.A. 1988 *Integer and Combinatorial Optimization*. Wiley, New York.
20. Rawls, J. 1972 *A Theory of Justice*, Oxford University Press, Oxford.
21. Rhodes, F. and Williams, H.P. 1995 Discrete subadditive functions as Gomory functions. *Mathematical Proceedings of the Cambridge Philosophical Society* **117** 559-574
22. Rödseth, O.J. 1978 On a linear Diophantine problem of Frobenius, *J. reine angew. Math.* **301** 171-178
23. Ryan, J. and Trotter, L.E. Jr 1994 Weyl-Minkowski duality for integral monoids, *Optimization* **28** 315-331
24. Scarf, H.E. 1977 An observation on the structure of production sets with indivisibilities *Proceedings of the National Academy of Sciences of the United States of America* **74** 3637-3641
25. Schrijver, A. 1986 *Theory of Linear and Integer Programming*. Wiley, New York.
26. Williams, H.P. 1984 A duality theorem for linear congruences. *Discrete Applied Mathematics* **7** 93-103
27. Williams, H.P. 1996 Constructing the value function for an integer linear programme over a cone. *Computational Optimisation and Applications* **6**, 15-26
28. Williams, H.P. 1996 The equivalence of two theorems of integer programming *Bulletin of the London Mathematical Society* **28** 311-316

29. Williams, H.P. 1997 Integer programming and pricing revisited. *IMA Journal of Mathematics Applied in Business and Industry* **8**, 203-214
30. Williams, H.P. 1984 A duality theorem for linear congruences. *Discrete Applied Mathematics* **7** 93-103
31. Williams, H.P. 1986 Fourier's method of linear programming and its dual. *The American Mathematical Monthly* **93** 681-695
32. Williams, H.P. 1999 *Model Building in Mathematical Programming*, 4th Edition, Wiley, Chichester.
33. Williams, H.P. 2009 *Logic and Integer Programming*, Springer, New York. Constructing the value function for an integer linear programme over a cone. *Computational Optimisation and Applications* **6**, 15-26
34. Wolsey, L.A. 1981 The b-hull of an integer programme. *Discrete Applied Mathematics* **3** 193-201