stichting

mathematisch

centrum

$\sum$ MC

JAC.M. ANTHONISSE
A NOTE ON EQUIVALENT SYSTEMS OF
LINEAR DIOPHANTINE EQUATIONS

BA

2e boerhaavestraat 49 amsterdam

Abstract

Another constructive proof is presented for the fact that a system of
linear equations with integer coefficients in bounded integer
variables is equivalent to a single equation, which is a linear
combination of the original ones. The equation is obtained in a number
of steps; in each step two equations are replaced by a single one.
This replacement is performed subject to the condition that the
remaining equations hold and a final equation with relatively small
coefficients is obtained. It may be inefficient however to calculate
small coefficients, as the original coefficients can be used to
represent the final ones in a suitably chosen number system.

Contents

1. Introduction

In their paper [2] Elmaghraby and Wig used two theorems, due to
Mathews [3], to aggregate a system of two linear equations with integer
coefficients in bounded, integer variables, into a single, equivalent
linear equation. By repeated application a system of m such equations
can be reduced to a single equation which is a linear combination of
the original ones.
The coefficients in the final equation however, tend to be rather
large.
In the preliminary paper [1] it was shown that smaller coefficients can
be obtained and that the original coefficients can be used to represent
the final coefficients in a suitably choosen number system.
In [4] Padberg derived some sharper results for the case of two equations.

In the present paper the results of [4] are improved and combined with
those of [1].

2. The Case m = 2

The aggregation of a system into a single equation is based upon the following, and well-known, result.

Theorem 1

If q_1 and q_2 are two relative prime integers, then all integer solutions of the equation

$$q_1 \, y_1 + q_2 \, y_2 = 0 \tag{1}$$

are of the form

$$\left.\begin{aligned} y_1 &= t \, q_2 \\[2ex] y_2 &= -t \, q_1, \end{aligned}\right\} \tag{2}$$

where t is any integer.

Proof

The equation (1) yields $y_1 = -\dfrac{q_2}{q_1} \, y_2$. As y_1 is required to be an integer, and the greatest common divisor of q_1 and q_2 is 1, y_2 must be a multiple of q_1. This completes the proof.

Now consider the equations

$$y_i = \sum_{j=1}^{n} a_{ij} \, x_j - a_{i0} = 0 \qquad (i=1,2), \tag{3}$$

$$\left.\begin{aligned} 0 &\le x_j \le b_j \\[2ex] x_j &= \text{integer} \end{aligned}\right\} \qquad (j=1,\ldots,n), \tag{4}$$

where a_{ij} and b_j are assumed to be integers. Consequently, y_i is integer valued.

It is easily seen that

$$L_i \leq y_i \leq U_i \qquad (i=1,2) \qquad (5)$$

where

$$L_i = \sum_{\substack{j=1 \\ a_{ij}<0}}^{n} a_{ij} b_j - a_{i0} \qquad (i=1,2) \qquad (6)$$

and

$$U_i = \sum_{\substack{j=1 \\ a_{ij}>0}}^{n} a_{ij} b_j - a_{i0} \qquad (i=1,2). \qquad (7)$$

If $L_i = 0$ the equation $y_i = 0$ implies $x_j = b_j$ if $a_{ij} < 0$ and $x_j = 0$ if $a_{ij} > 0$. In this case these substitutions can be performed and only a single equation, in fewer variables remains.

A similar result holds if $U_i = 0$.

If $L_i > 0$ or $U_i < 0$ the equation $y_i = 0$ has no solution and the system is infeasible.

Define

$$S_0 = \{(u_1,u_2) \mid L_i \leq u_i \leq U_i, \ u_i = \text{integer} \ (i=1,2)\}. \qquad (8)$$

Theorem 2

For any two relative prime integers q_2 and q_1 such that

$$(q_2,-q_1) \notin S_0$$

and

$$(-q_2,q_1) \notin S_0, \qquad \left.\begin{array}{c} \\ \\ \\ \\ \\ \end{array}\right\} \qquad (9)$$

the unique solution of

$$q_1 y_1 + q_2 y_2 = 0 \qquad (1)$$

$$\left.\begin{array}{c} L_i \le y_i \le U_i \\ \\ y_i = \text{integer} \end{array}\right\} \qquad (i=1,2) \qquad (10)$$

is $y_1 = y_2 = 0$.

Proof

Each solution of (1) is of the form $y_1 = tq_2$, $y_2 = -tq_1$.

If $t \ne 0$ then (9) leads to the conclusion that $(y_1, y_2) \notin S_0$, contradicting (10).

This completes the proof.

It is easily seen that system (3), (4) is equivalent to:

$$\sum_{j=1}^{n} (q_1 a_{1j} + q_2 a_{2j}) x_j - (q_1 a_{10} + q_2 a_{20}) = 0 \qquad (11)$$

$$\left.\begin{array}{c} 0 \le x_j \le b_j \\ \\ x_j = \text{integer} \end{array}\right\} \qquad (j=1,\ldots,n), \qquad (4)$$

for any two relative prime integers satisfying (9). Valid choices are:

$$q_1 \ge U_2 + 1 \qquad \text{and} \quad q_2 \ge U_1 + 1,$$

$$\text{or} \quad q_1 \ge -L_2 + 1 \qquad \text{and} \quad q_2 \ge -L_1 + 1,$$

$$\text{or} \quad q_1 \ge 1 \qquad \text{and} \quad q_2 \ge 1 + \max(U_1, -L_1),$$

$$\text{or} \quad q_1 \ge 1 + \max(U_2, -L_2) \quad \text{and} \quad q_2 \ge 1.$$

The above results were given by Padberg [3].

It should be noted that theorem 2 is based upon the integrality of y_i. The fact that y_i represents a linear function will be exploited to obtain smaller $|q_i|$.

Define $v(p)$ = the minimum and $w(p)$ = the maximum of

$$\sum_{j=1}^{n} a_{2j} x_j - a_{20} \tag{12}$$

subject to

$$\left. \begin{array}{l} \sum_{j=1}^{n} a_{1j} x_j - a_{10} = p \\ \\ 0 \leq x_j \leq b_j \qquad\qquad (j=1,\ldots,n), \end{array} \right\} \tag{13}$$

for all integers p such that

$$L_1 \leq p \leq U_1.$$

Both $v(p)$ and $w(p)$ are piece-wise linear functions, $v(p)$ is convex, $w(p)$ is concave.

Define furthermore,

$$S_1 = \{(u_1,u_2) \mid L_1 \leq u_1 \leq U_1, \ v(u_1) \leq u_2 \leq w(u_1), \ u_i = \text{integer}\}. \tag{14}$$

Theorem 3

For any two relative prime integers q_2 and q_1 such that

$$(q_2,-q_1) \notin S_1 \quad \text{and} \quad (-q_2,q_1) \notin S_1, \tag{15}$$

system (3), (4) is equivalent to system (11), (4).

Proof

As (11) is a lineair combination of (3) each solution of (3), (4) obviously solves (11), (4). Assume $x_j = \underline{x}_j$ (j=1,...,n) satisfies (11), (4) but not (3).

Define

$$\underline{y}_i = \sum_{j=1}^{n} a_{ij} \underline{x}_j - a_{i0} \qquad (i=1,2), \qquad (16)$$

then (11) yields

$$q_1 \underline{y}_1 + q_2 \underline{y}_2 = 0,$$

so $\underline{y}_1 = tq_2$ and $\underline{y}_2 = -tq_1$, with $t \neq 0$.

This implies that $(\underline{y}_1, \underline{y}_2) \notin S_1$.

Both relations $\underline{y}_1 < L_1$ and $\underline{y}_1 > U_1$ contradict (4).

If $L_1 \leq \underline{y}_1 \leq U_1$ then either $\underline{y}_2 < v(\underline{y}_1)$ or $\underline{y}_2 > w(\underline{y}_1)$, contradicting the definition of $v(p)$ or $w(p)$ respectively.

This completes the proof.

As $S_1 \subset S_0$ theorem 3 may lead to smaller $|q_i|$ than can be obtained by theorem 2.

It is not difficult to determine the functions $v(p)$ and $w(p)$.
Consider the example from [3].

$$y_2 = -x_1 + 3x_2 - 5x_3 - x_4 + 4x_5 + x_6 + 2 = 0$$

$$y_1 = 2x_1 - 6x_2 + 3x_3 + 2x_4 - 2x_5 + x_7 \qquad = 0$$

$$x_i \in \{0,1\} \qquad (i=1,\dots,5)$$

$$0 \leq x_6 \leq 5, \quad 0 \leq x_7 \leq 8, \quad \text{integer}$$

First substitute $x_2 = 1 - x_2'$, $x_5 = 1 - x_5'$, then rearrange the variables in such a way that the ratio a_{2j}/a_{1j} is non-increasing $(j=1,\dots,n)$.

$$y_2 = x_6 \qquad - x_1 - 3x_2' - x_4 - 5x_3 - 4x_5' + 9 = 0$$

$$y_1 = \qquad x_7 + 2x_1 + 6x_2' + 2x_4 + 3x_3 + 2x_5' - 8 = 0$$

$v(p)$ is found by working from the right to the left:

p		v(p)	
-8	= -8	+9	= +9
-8 + 2	= -6	+9 - 4	= +5
-8 + 2 + 3	= -3	+9 - 4 - 5	= 0
-8 + 2 + 3 + 2	= -1	+9 - 4 - 5 - 1	= -1
-8 + 2 + 3 + 2 + 6	= +5	+9 - 4 - 5 - 1 - 3	= -4
-8 + 2 + 3 + 2 + 6 + 2	= +7	+9 - 4 - 5 - 1 - 3 - 1 = -5	
-8 + 2 + 3 + 2 + 6 + 2 + 8 = +15		+9 - 4 - 5 - 1 - 3 - 1 = -5	

Similarly, $w(p)$ is found by working in the opposite direction:

p		w(p)	
-8	= -8	+9 + 5	= +14
-8 + 8	= 0	+14 + 0	= +14
0 + 2	= +2	+14 - 1	= +13
+2 + 6	= +8	+13 - 3	= +10
+8 + 2	= +10	+10 - 1	= +9
+10 + 3	= +13	+9 - 5	= +4
+13 + 2	= +15	+4 - 4	= 0

The situation is depicted in figure 1, where the functions $v(p)$ and $w(p)$ are also drawn for non-integer p, and $-v(-p)$ is given if $v(p) < 0$. Evidently, $(-7,6) \notin S_1$, and $(7,-6) \notin S_1$, the choice $q_2 = 7$, $q_1 = 6$ leads to:

$$7x_6 + 6x_7 + 5x_1 + 15x_2' + 5x_4 - 17x_3 - 16x_5' + 15 = 0,$$

or, with $x_3 = 1 - x_3'$, $x_5' = 1 - x_5$,

$$7x_6 + 6x_7 + 5x_1 + 15x_2' + 5x_4 + 17x_3' + 16x_5 = 18.$$

$q_2 = 11$, $q_1 = 6$ yields:

$$11x_6 + 6x_7 + x_1 + 3x_2' + x_4 + 37x_3' + 32x_5 = 18,$$

so $x_3 = 1$ and $x_5 = 0$ in any solution.

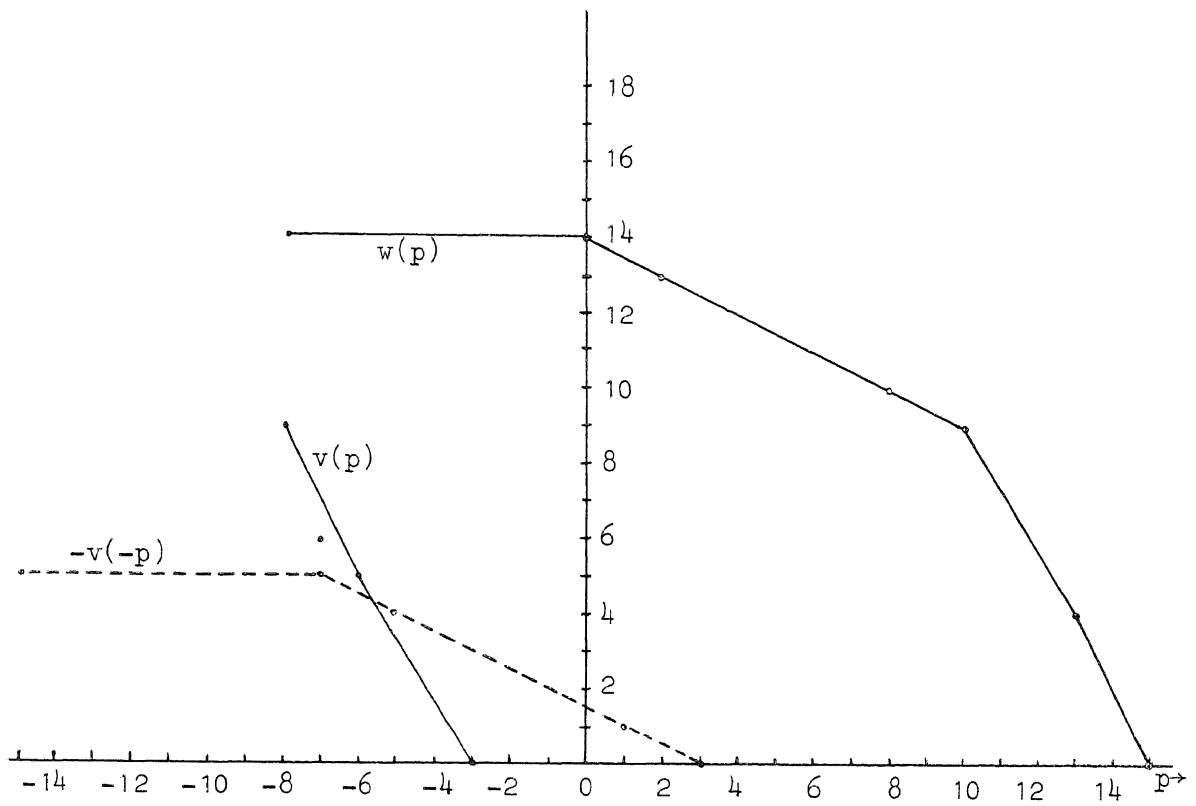This leads to the conclusion that it may be worthwhile to select other than the 'minimal' q_2 and q_1.

figure 1

3. The General Linear Case

A system of m equations

$$y_i = \sum_{j=1}^{n} a_{ij} \, x_j - a_{i0} = 0 \qquad (i=1,\ldots,m), \qquad (17)$$

$$\left. \begin{array}{l} 0 \le x_j \le b_j \\[2ex] x_j = \text{integer} \end{array} \right\} \qquad (j=1,\ldots,n), \qquad (4)$$

where b_j and a_{ij} denote integers, can be aggregated into a single, equivalent equation

$$\sum_{j=1}^{n} a_j \, x_j - a_0 = 0, \qquad (18)$$

$$\left. \begin{array}{l} 0 \le x_j \le b_j \\[2ex] x_j = \text{integer} \end{array} \right\} \qquad (j=1,\ldots,n), \qquad (4)$$

by $m-1$ applications of theorem 3. Each coefficient a_j $(j=0,1,\ldots,n)$ in (18) is a linear combination of the coefficients $a_{1j},\ldots,a_{mj}$. The $|a_j|$ may be very large. The next theorem leads to smaller coefficients.

Define L_{m-1} = the minimum, and U_{m-1} = the maximum of

$$y_{m-1} \qquad (19)$$

subject to

$$\left. \begin{array}{ll} y_i = 0 & (i=1,\ldots,m-2), \\[2ex] 0 \le x_j \le b_j & (j=1,\ldots,n), \end{array} \right\} \qquad (20)$$

where y_i denotes the function $\sum_{j=1}^{n} a_{ij} \, x_j - a_{i0}$.

Define $v(p)$ = the minimum, and $w(p)$ = the maximum of

$$y_m \qquad\qquad (21)$$

subject to

$$\left. \begin{array}{l} y_i = 0 \qquad (i=1,\ldots,m-2) \\[1em] y_{m-1} = p \end{array} \right\} \qquad (22)$$

for all integers p such that

$$L_{m-1} \le p \le U_{m-1}.$$

Again, both $v(p)$ and $w(p)$ are piece-wise linear, $v(p)$ is convex, $w(p)$ is concave.

Finally, define

$$S_2 = \{(u_1,u_2) \mid L_{m-1} \le u_1 \le U_{m-1}, \; v(u_1) \le u_2 \le w(u_1), \; u_i = \text{integer}\}. \quad (23)$$

<u>Theorem 4</u>

For any two relative prime integers q_m and q_{m-1} such that

$$(q_m,-q_{m-1}) \notin S_2 \quad \text{and} \quad (-q_m,q_{m-1}) \notin S_2, \qquad (24)$$

system (17), (4) is equivalent to

$$\left. \begin{array}{l} y_i = 0 \qquad (i=1,\ldots,m-2) \\[1em] q_{m-1}\,y_{m-1} + q_m\,y_m = 0 \end{array} \right\} \qquad (25)$$

$$\left. \begin{array}{l} 0 \le x_j \le b_j \\[1em] x_j = \text{integer} \end{array} \right\} \qquad (j=1,\ldots,n). \qquad (4)$$

Proof

Assume $x_j = \underline{x}_j$ solves (25), (4), but not (17). Then

$$\underline{y}_i = \sum_{j=1}^{n} a_{ij} \underline{x}_j - a_{i0} \qquad (i=m-1,m)$$

satisfy $\underline{y}_{m-1} = tq_m$ and $\underline{y}_m = -tq_{m-1}$, with $t \neq 0$.

This implies $(\underline{y}_{m-1}, \underline{y}_m) \notin S_2$.

But $\underline{x}_j$ satisfies (25), so $L_{m-1} \leq \underline{y}_{m-1} \leq U_{m-1}$ and $v(\underline{y}_{m-1}) \leq \underline{y}_m \leq w(\underline{y}_{m-1})$ hold by the definitions of L_{m-1}, U_{m-1} and $v(p)$, $w(p)$ respectively.

This contradiction completes the proof.

As $S_2 \subset S_1$ the resulting q_{m-1} and q_m may be, in absolute value, much smaller than those obtained by the application of theorem 3 to the equations $y_{m-1} = 0$, $y_m = 0$.

Theorem 3 can be used with S_2 replaced by

$$S_3 = \{(u_1, u_2) \mid L_{m-1} \leq u_1 \leq U_{m-1}, \; L_m \leq u_2 \leq U_m, \; u_i = \text{integer}\},$$

where L_m and U_m denote the minimum and maximum of y_m subject to (20), respectively.

Consider the problem

$$
\begin{aligned}
y_1 &= \quad\quad\; x_2 - 2x_3 + x_4 + x_5 + x_8 + 1 = 0 \\
y_2 &= 2x_1 - 6x_2 + 3x_3 + 2x_4 - 2x_5 + x_7 \quad\;\; = 0 \\
y_3 &= -x_1 + 3x_2 - 5x_3 - x_4 + 4x_5 + x_6 + 2 = 0 \\
&\quad\; x_i \in \{0,1\} \qquad (i=1,\ldots,5) \\
&\quad\; 0 \leq x_6 \leq 5, \;\; 0 \leq x_7 \leq 8, \;\; 0 \leq x_8 \leq 1, \;\; \text{integer}
\end{aligned}
$$

which is the previous example with an additional constraint. This example was used in [1]. $y_1 = 0$ implies $x_3 = 1$, after this substitution it is seen that $x_2 = 1$ in any solution, this substitution yields $x_4 = x_5 = x_8 = 0$. The only solutions are:

$$x_1 = x_6 = 0, \quad x_7 = 3 \quad \text{and} \quad x_1 = x_6 = x_7 = 1.$$

Now the system will be aggregated into a single equation.

$y_1 = 0$ implies $-3 \leq y_2 \leq 15$.

Figure 2 contains the $-v(-p)$ and $w(p)$ of y_3 subject to $y_1 = 0$, $y_2 = p$ $(p=-3,\ldots,15)$. As $(-4,5) \notin S_2$ and $(4,-5) \notin S_2$ the system is equivalent to

$$y_1 = 0$$

$$5y_2 + 4y_3 = 0,$$

or

$$y_1 = \quad x_2 + 2x_3' + x_4 + x_5 \qquad + x_8 - 1 = 0$$

$$y_4 = 6x_1 - 18x_2 + 5x_3' + 6x_4 + 6x_5 + 4x_6 + 5x_7 \qquad + 3 = 0$$

Minimizing and maximazing y_4 subject to $y_1 = p$ yields

p	v(p)	w(p)
-1	+3	+69
0	-15	+75
1	-15	+81
2	-12.5	+83.5
3	-10	+86
4	-4	+86
5	+2	+68 ,

see figure 3.

Now $(-5,1) \notin S_1$ and $(5,-1) \notin S_1$, leading to $y_1 + 5y_4 = 0$ or:

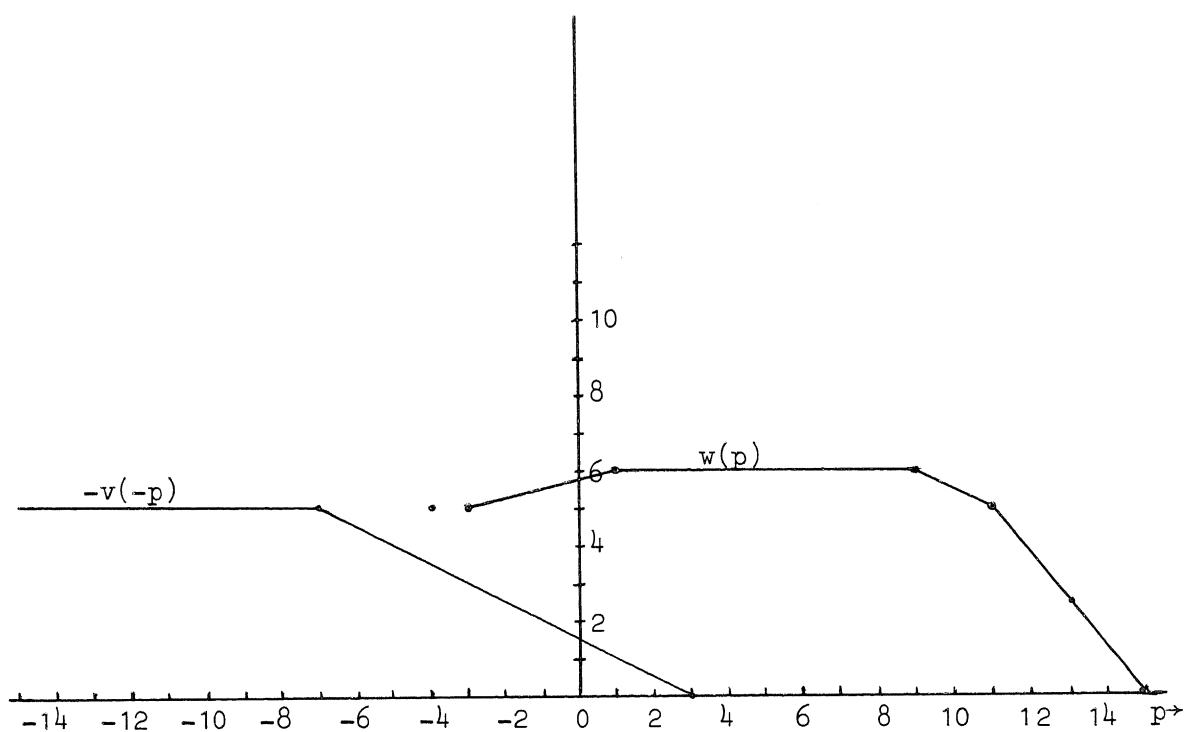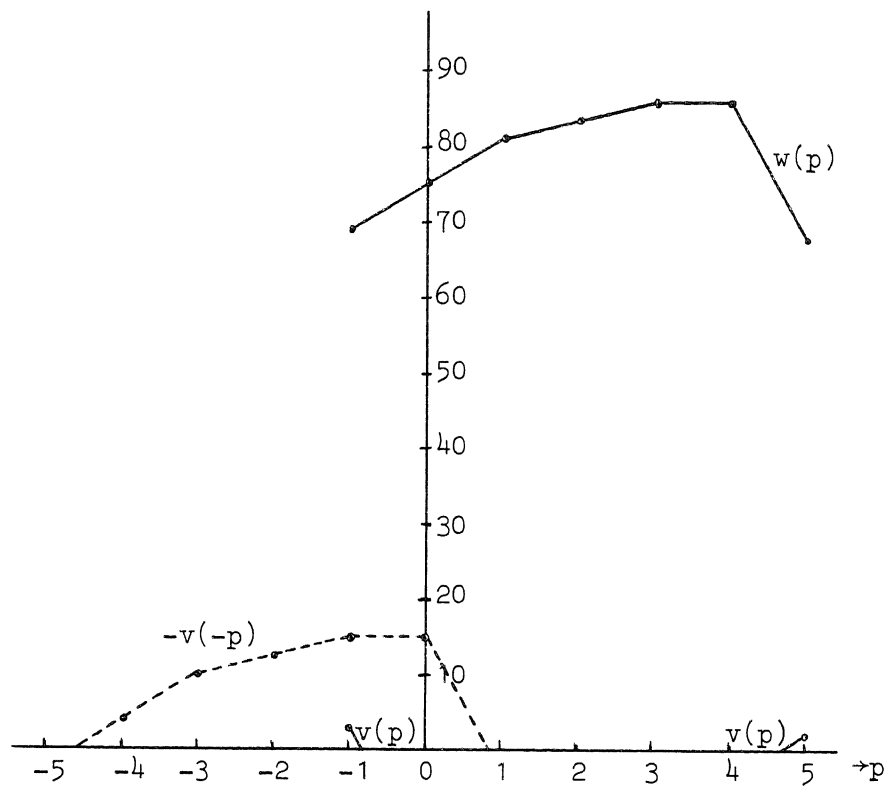$$30x_1 + 89x_2' + 27x_3' + 31x_4 + 20x_6 + 25x_7 + x_8 = 75.$$

figure 2

figure 3

4. A Generalisation

Part of the previous results can be obtained without using the integrality of the functions y_i. Let y_1 and y_2 denote real valued functions, which are defined over an arbitrary domain.

Theorem 5

If the function y_1 is bounded, i.e. $|y_1| \leq B$, and $y_2 \neq 0$ implies $|y_2| \geq \varepsilon > 0$ then, for any q_1 satisfying $|q_1| > B/\varepsilon$, the system of equations

$$\left. \begin{array}{l} y_1 = 0 \\ \\ y_2 = 0 \end{array} \right\} \tag{26}$$

is equivalent to

$$y_1 + q_1 y_2 = 0 \tag{27}$$

Proof

Obviously, any solution of (26) satisfies (27). If (27) holds and $y_2 = 0$ then $y_1 = 0$, if $y_2 \neq 0$ then

$$|y_1| = |q_1 y_2| \geq |q_1| \varepsilon > B.$$

This completes the proof.

Consider the system of equations

$$y_i = 0 \qquad (i=1,\ldots,m), \tag{28}$$

where $y_i \neq 0$ implies $|y_i| \geq 1$ $(i=2,\ldots,m)$.

Define

$$c_i = \sup(|y_i| \mid y_k = 0 \ (k=1,\ldots,i-1)), \qquad (i=1,\ldots,m-1). \tag{29}$$

Theorem 6

For any q_i satisfying $|q_i| \geq C_i + 1$ the system (28) is equivalent to

$$\sum_{i=1}^{m} q_1 \, q_2 \cdots q_{i-1} \, y_i = 0.$$

Proof

It is easily seen that system (28) is equivalent to

$$y_i = 0 \qquad\qquad (i=1,\ldots,m-2)$$

$$y_{m-1} + q_{m-1} \, y_m = 0.$$

The first $m-2$ equations imply $|y_{m-1}| < |q_{m-1}|$. If $y_m \neq 0$ the last one implies

$$|y_{m-1}| = |q_{m-1}| \cdot |y_m| \geq |q_{m-1}|.$$

Now assume $m > 2$.

If $y_{m-1} = 0$ or $y_m = 0$ then $y' = y_{m-1} + q_{m-1} \, y_m \neq 0$ implies $|y'| \geq 1$. If $y_{m-1} \neq 0$ and $y_m \neq 0$ then $|y_{m-1}| \leq C_{m-1}$ and $|q_{m-1} \, y_m| \geq C_{m-1} + 1$ yield the same implication.

This completes the proof.

The above theorem leads to the conclusion that y_1 should be bounded, y_2 should be bounded on that part of the domain where $y_1 = 0$, y_3 bounded on that part where $y_1 = y_2 = 0$ and so on. y_m, however, may be unbounded.

During the computation of C_i it might be found that $y_k = 0$ (k=1,...,i-1) implies $y_i \neq 0$. In this case the system is infeasible.
If $C_i = 0$ the equation $f_i = 0$ is redundant.

5. Numerical Aspects

This discussion is restricted to the linear case. It is easily seen
that the aggregation of a system of equations may lead to rather large
coefficients in the final equation. The coefficients can be decreased by
using small q_i but these can be obtained at a rather high computational
price only.

If theorem 6 is used, the coefficients of the original system can be
transformed into a representation of the final coefficients.

The original system is (18), (4). Let q_i (i=1,...,m-1) integers satis-
fying

$$q_i \geq 1 + \max(|y_i| \mid y_k = 0 \ (k=1,...,i-1), \ 0 \leq x_j \leq b_j, \ (j=1,...,n)).$$

Then the system is equivalent to

$$\sum_{j=1}^{n} a_j x_j = a_0$$

$$\left.\begin{array}{l} 0 \leq x_j \leq b_j \\ \\ x_j = \text{integer} \end{array}\right\} \qquad (j=1,...,n),$$

where

$$a_j = \sum_{i=1}^{m} q_1 \cdots q_{i-1} a_{ij} \qquad (j=0,1,...,n).$$

If $0 \leq a_{ij} < q_i$ (i=1,...,m-1) and $a_{mj} \geq 0$ then the j-th column
$(a_{1j},...,a_{mj})$ from the matrix (a_{ij}) can be interpreted as the represen-
tation of a_j in a, possibly unfamiliar, number system determined by the
q_i.
This number system has q_i 'digits' from the i-th position (i=1,...,m-1),
the number of digits for the m-th position is unbounded.

If $-q_i < a_{ij} \leq 0$ $(i=1,\ldots,m-1)$ and $a_{mj} \leq 0$ then $(a_{1j},\ldots,a_{mj})$ represents a_j in the same system.

With the convention that all 'digits' are either non-negative or non-positive any integer has a unique representation in the system.

Thus a_j can be computed by transforming the column $(a_{1j},\ldots,a_{mj})$ into the representation of a_j. This is not difficult as

$$q_1 \cdots q_{i-1} a_{ij} + q_1 \cdots q_i a_{i+1,j} =$$

$$= q_1 \cdots q_{i-1}(a_{ij}+q_i) + q_1 \cdots q_i(a_{i+1,j}-1) =$$

$$= q_1 \cdots q_{i-1}(a_{ij}-q_i) + q_1 \cdots q_i(a_{i+1,j}+1).$$

6. References

[1] Jac.M. Anthonisse

A Note on Reducing a System to a Single Equation,

Mathematisch Centrum, Amsterdam,

preliminary report BN 1/70, December 1970.

[2] S.E. Elmaghraby and M.K. Wig

On the treatment of stock cutting problems as diophantine programs,

North Carolina State University and Corning Glass Research Center,

Report No. 61, May 11, 1970.

[3] G.B. Mathews

On the partition of numbers,

Proceedings of the London Mathematical Society,

Vol. 28, 1897, pp. 486-490.

[4] M.W. Padberg

Equivalent Knapsack-type Formulations of Bounded Integer Linear

Programs,

Carnegie-Mellon University,

Management Sciences Research Report No. 227, September 1970.