Column Basis Reduction and Hard Knapsack Problems

Bala Krishnamoorthy

joint work with
Gábor Pataki

University of North Carolina at Chapel Hill

Feb 26, 2004

Integer Programming (IP)

- The integer programming problem - feasibility version

$$\text{Given} \quad A_{m\times n},\ d_{m\times 1},\ \text{rational}$$

$$\text{Find } x \in \mathcal{Z}^n,\ Ax \le d \qquad \text{or prove that } \nexists x$$

- The integer programming problem - optimization version

$$\begin{aligned} \max \quad & c^T x & & (c_{n\times 1}) \\ \text{subject to} \quad & Ax & \le d & \\ & x & \in \mathcal{Z}^n & \end{aligned}$$

- two versions equivalent

Integer Programming (IP)

- wide variety of applications
  - airline crew scheduling, facility location
  - machine learning and statistical classification
  - side-chain structure prediction (proteins)
- IP (in arbitrary dimensions) is NP-complete
- several efficient solution methods
  (branch-and-bound, cutting planes)
- *small* problems - intractable to solvers

Example - *hard* problem

$$\begin{aligned} \text{max} \quad & 12228\,x_1 + 36679\,x_2 + 36682\,x_3 \\ & + \; 48908\,x_4 + 61139\,x_5 + 73365\,x_6 \end{aligned}$$

$$\begin{aligned} \text{subject to } & 12228\,x_1 + 36679\,x_2 + 36682\,x_3 \\ & + \; 48908\,x_4 + 61139\,x_5 + 73365\,x_6 \; \le 89716837 \\ & x_i \ge 0, \quad x_i \in \mathcal{Z} \quad \forall i \end{aligned}$$

from Wolsey: Integer Programming (1999).

CPLEX takes more than 400 million B&B nodes!

Basis reduction

- Given integral matrix A, basis reduction (BR) computes a unimodular $U (\Leftrightarrow \det U = \pm 1)$ st. the columns of AU are "short" and "nearly" orthogonal.

  Example

$$A = \begin{pmatrix} 289 & 18 \\ 466 & 29 \\ 273 & 17 \end{pmatrix}, \; U = \begin{pmatrix} 1 & -15 \\ -16 & 241 \end{pmatrix}, \; AU = \begin{pmatrix} 1 & 3 \\ 2 & -1 \\ 1 & 2 \end{pmatrix}$$

- Computing $AU \Leftrightarrow$ doing *elementary column operations* on A: - adding an integer multiple of a column to another; multiplying a column by -1; swapping columns.

- A "good" U can be computed in polynomial time (Lovász; Kannan; Schnorr, . . .).

The use of BR in integer programming

- H. W. Lenstra (1983); Kannan (1987): polytime algorithms for IP in fixed dimensions.

- Lagarias, Odlyzko (1985): Solving low-density subset sum problems. . . .

- Aardal, Hurkens, A.K. Lenstra (1998),
  Aardal, Bixby, Hurkens, A.K. Lenstra, J.W. Smeltink (1999),
  Louvaux, Wolsey (2001):
  Reformulation of hard, equality-constrained IPs.

Reformulating equality-constrained feasibility IPs

From A, H, L (1998); A, B, H, L, S (1999).
We need to find

$$x \in \mathcal{Z}^n,\ \ell \leq x \leq u,\ Ax = d$$

(all data integer, $A_{m \times n},\ m \leq n$), or prove that $\nexists x$.

Full-dimensional Reformulation:

$$Ax = b, x \in \mathcal{Z}^n \Leftrightarrow x = B\lambda + x_d : \text{for some } \lambda \in \mathcal{Z}^{n-m}$$

where x_d and B satisfy $Ax_d = d$, $AB = 0$.

Full-dimensional Reformulation

- B and x_d are integral.
- columns of $[B, x_d]$ short and nearly orthogonal.
- found by doing BR on an enlarged matrix using two large constants N_1, N_2.

Technique similar to the one in Lagarias/Odlyzko (1985), and later papers on subset sum.

- The reformulated problem of finding

$$\lambda \in \mathcal{Z}^{n-m},\ \ell \leq B\lambda + x_d \leq u$$

proved *much* easier to solve for some problems (e.g. marketshare of Cornuéjols & Dawande).

Analysis for a special IP (A & L)

Aardal & Lenstra (2002): If

- $[A, d] = [a^T, a_0]$ (i.e. $m = 1$), $\ell = 0$, $u = +\infty$ and
- $a^T = p^T M + r^T$ with M large compared to p, r,

then

- $\exists\, a_0 \geq \text{const} * M^2$, so problem is infeasible.
- In the reformulation, the length of last column of B is lower bounded $\Rightarrow$ it is intuitively good to branch on λ_{n-1}.

Main questions

1. Do we have to change the dimension of the problem (go from $\mathcal{Z}^n$ to $\mathcal{Z}^{n-m}$) ?

2. Tighter analysis (going beyond lower bd on last vector's length) ?

3. Analysis and computation for more general problem classes?

Rest of the talk

1. Simplified reformulation for arbitrary IPs:
   range space / null space reformulations.

2. Decomposable knapsack problem with inequality/equality constraints. Includes

   - A & L's equality constrained knapsack problem
   - some of Chvátal's hard knapsack problems:
     Chvátal-Todd, Avis; (1980)
   - $t + 1$-level problems: decomposition into $t + 1$ vectors, not just 2.

3. These problems are

   - Provably hard for B&B branching on x_i variables.
   - Provably easy both for range space, or null space reformulations!

Range Space Reformulation

$$\begin{aligned} P &= \{\, x \,|\, \ell \leq \ Ax \leq b \,\} \\ \tilde{P} &= \{\, y \,|\, \ell \leq (AU)y \leq b\} \end{aligned}$$

where U is unimodular

There is 1-1 correspondence between

$$P \cap \mathcal{Z}^n \text{ and } \tilde{P} \cap \mathcal{Z}^n$$

given by

$$Uy \ = \ x$$

- Choose U so columns of AU are reduced.
- can be applied when some (or all) of the "$\leq$" are "$=$".

Null Space reformulation

If

$$A_1 x \;=\; b_1$$

is a subset of the inequalities in $\ell \leq Ax \leq b$, then

$$A_1 x = b_1, x \in \mathcal{Z}^n \Leftrightarrow x = B_1 \lambda + x_d : \text{for some } \lambda \in \mathcal{Z}^{n-m_1}$$

where x_d and B_1 satisfy $A_1 x_d = b_1$, $A_1 B_1 = 0$.

- $[B_1, x_d]$ found by a Hermite Normal Form (HNF) computation;
  columns are *not* in general short and orthogonal.

- Substitute $B_1\lambda + x_d$ for x, and do the range space reformulation.

2-level decomposable knapsack problems

Knapsack feasibility problem:

$$(KP_2) \quad lb \le a^T x \le ub, \qquad 0 \le x \le u, \quad x \in \mathcal{Z}^n,$$

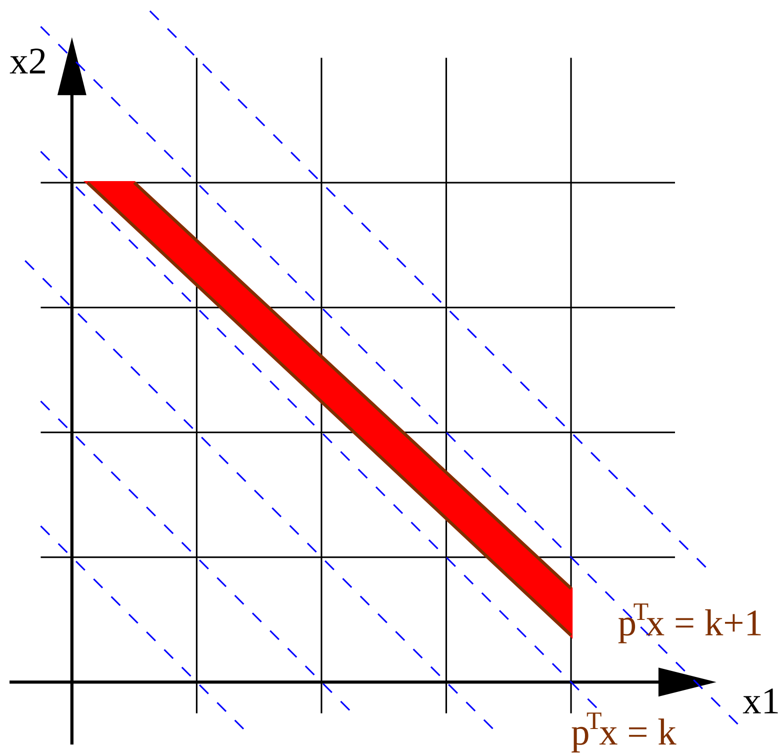
where

- $a = pM + r$, with p, $r \in \mathcal{Z}^n$; M large compared to p, r;
- lb, ub appropriately chosen, so KP_2 is LP-feasible, but integer infeasible.

A family of hard, infeasible knapsack problems

For some integer k, the feasible set of the LP relaxation is sandwiched between

$$p^T x = k \;\; and \;\; p^T x = k + 1.$$



A family of hard, infeasible knapsack problems

- Hardness for B&B using x_js for branching: if
  - $p^T u$ even, $k = p^T u/2$, and $p_1 u_1 = \cdots = p_n u_n$, $\Rightarrow$ no nodes are pruned above level $n/2$.
  - $u = +\infty$, $lb = ub$ appropriate $\Rightarrow$ # of B&B nodes is $\geq const * M^{n-1}$.

- Branching on $p^T x$ proves infeasibility in one step!

Basis reduction in range space

Recall general reformulation:

$$P = \{x \,|\, \ell \leq Ax \leq b\} \Leftrightarrow \tilde{P} = \{y \,|\, \ell \leq \ (AU)y \leq b\}$$

Basis reduction in range space

We use the reformulation with

$$A = \begin{pmatrix} a^T \\ I \end{pmatrix} = \begin{pmatrix} p^T M + r^T \\ I \end{pmatrix},$$

U is unimodular st. $AU = \begin{pmatrix} (p^T U)M + r^T U \\ U \end{pmatrix}$ is reduced.

Theorem: M suff. large $\Rightarrow$

$$p^T U = (\,\overbrace{0 \;\ldots\; 0}^{n-1}\;\alpha\,) \quad \text{for some } \alpha \in \mathcal{Z} \setminus \{\,0\,\}.$$

Corollary:

$$U\,y = x \;\Rightarrow\; p^T U y \;=\; p^T x \;\;\Rightarrow\;\; \alpha y_n = \; p^T x$$

$\Rightarrow$ branching on y_n in reformulation proves infeasibility in 1 step!

Basis reduction in null space

$$(KP_2) \quad lb \le a^T x \le ub, \quad 0 \le x \le u, \quad x \in \mathcal{Z}^n$$

Can be used if $lb = ub \Rightarrow$ reformulation has $n-1$ variables.

We can similarly prove

Theorem: M large $\Rightarrow \alpha y_{n-1} = p^T x +$ const.

Corollary: $\Rightarrow$ branching on y_{n-1} proves infeasibility in 1 step!

$(t+1)$-level decomposable knapsack problems

- For $a = p_1 M_1 + p_2 M_2 + \cdots + p_t M_t + p_{t+1}$, with $M_1 \gg M_2 \gg \cdots \gg M_t$ and appropriate ub, lb

$$(KP_{t+1}) \quad lb \leq a^T x \leq ub, \quad 0 \leq x \leq u, \quad x \in \mathcal{Z}^n$$

Problem is

- hard, if branching on x_j variables.
- easy, if branching on $p_1^T x$, $p_2^T x$, . . . , $p_t^T x$. But: if we stop before $p_t^T x \Rightarrow$ remaining problem is still hard!

Range Space Reformulation

We use the range space reformulation with

$$A = \begin{pmatrix} a^T \\ I \end{pmatrix}, \text{ choosing } U \text{ unimodular st. } AU \text{ is reduced.}$$

Theorem: $M_1 \gg M_2 \gg \cdots \gg M_t \Rightarrow$

$$\begin{pmatrix} p_1^T \\ p_2^T \\ \vdots \\ p_t^T \end{pmatrix} U = \begin{pmatrix} 0 & 0 \ldots & 0 & 0 & 0 & * \\ 0 & 0 \ldots & 0 & 0 & * & * \\ \vdots & & & & & \\ 0 & 0 \ldots & * & \ldots & * & * \end{pmatrix}$$

Range Space Reformulation

Corollary: Branching on $y_n, y_{n-1}, \ldots, y_{n-t}$ in reformulation

$\Leftrightarrow$ branching on $p_1^T x$, $p_2^T x$, . . . , $p_t^T x$ in original problem.

Analogous result for null space reformulation.

This way the problem solves after examining $2t$ B&B nodes.

Range Space Reformulation

Remark: When computing U, we do not know that a decomposes!!

For *fixed* t, the size of $M_1, M_2, \ldots, M_t$ can be chosen to be polynomial in the size of $p_1, \ldots, p_t, p_{t+1}$.

Corollary of independent interest: The Hermite Normal Form of

$$\begin{pmatrix} p_1^T \\ p_2^T \\ \vdots \\ p_t^T \end{pmatrix}$$

(a transformation using integral eco's to bring it into reduced echelon form) can be computed in poly time only from the aggregated p_i's, i.e. $p_1 M_1 + p_2 M_2 + \cdots + p_t M_t$, if t is fixed.

Example 1: Chvátal-Todd knapsack problem

$$\delta \leq a^T x \leq \gamma; \quad 0 \leq x \leq e, \; x \in \mathcal{Z}^n; \text{ with}$$

$$a_i = 2^n + 2^{i-1}, \quad e = (1\,1 \ldots 1)^T$$

From Chvátal "Hard knapsack problems" (1980).

- Known: B&B takes a tree of depth at least $\lfloor \frac{n}{2} \rfloor$ to prove infeasibility.

- New:
  - Branching on $e^T x$ proves infeasibility in 1 step.
  - After BR (either in range space or in null space) branching on last variable proves infeasibility in 1 step. Reason: This is just a 2–level problem with $M = 2^n$, $r = (1 \ 2 \ \ldots \ 2^{n-1})$ and $p = e$.

Example 2

3-level 0–1 knapsack problem (generated by us):

- $n = 100$ variables, M_1, M_2 smaller than what theory requires $\Rightarrow$ $a_j \leq 8000$, lb, $ub \leq 100,000$.
- Original problem takes more than 2 billion B&B nodes for CPLEX.
- Reformulation solves at rootnode.

Computational results

Goal: Solve IPs that are hard for B&B , or branch-and-cut, have both equality and inequality constraints, and go beyond the classes analyzed.

We solve optimization problems as:

– as a sequence of feasibility problems via binary search.
– reformulating the objective c^T by doing BR on

$$\begin{pmatrix} c^T \\ A \end{pmatrix},$$

where A is the constraint matrix.
BR: by LiDIA at the Univ. of Darmstadt; NTL (V. Shoup)
IP solver: CPLEX 7.1; Machine: 450 MHz Sun.

Maximization versions of integer subset sum

First four: Cornuéjols, Urbaniak, Weismantel, Wolsey (1998).

Last one from Wolsey: Integer Programming (1999).

$$
\begin{array}{llllllllll}
ex_1: & 12346x_1 & +14666x_2 & +12366x_3 & +13466x_4 & +13446x_5 & +14566x_6 & \leq & 685111 \\
ex_2: & 12212x_1 & +12214x_2 & +12216x_3 & +24416x_4 & +24418x_5 & & \leq & 27123101 \\
ex_3: & 12223x_1 & +12224x_2 & +12225x_3 & +12226x_4 & +12227x_5 & +12228x_6 & \leq & 1123100 \\
ex_4: & 1124x_1 & +1366x_2 & +1566x_3 & +2566x_4 & +3566x_5 & +5566x_6 & \leq & 148801 \\
wol: & 12228x_1 & +36679x_2 & +36682x_3 & +48908x_4 & +61139x_5 & +73365x_6 & \leq & 89716837
\end{array}
$$

$x_i \geq 0 \; \forall i$

Number of B&B nodes after column BR: 16, 0, 2, 6, 4.

Strongly correlated knapsack problems

$$\begin{aligned} \max \quad & (a^T + ke^T)x \\ & a^T x \leq \beta \\ & x_i \in \{0,1\} \;\; i = 1, \ldots, n \end{aligned}$$

where

- $n = 100$, $a_i \in [R/10, R]$ random, uniform,
- $k = 10$, $\beta = \lfloor (je^T a)/5 \rfloor$ in instance j.
- $R = 10^5, 10^6, 10^7$.

Strongly correlated knapsack problems

	$R = 10^5$		$R = 10^6$		$R = 10^7$	
Inst	CPU	#BB	CPU	#BB	CPU	#BB
1	2.3	1231	2.7	1206	2680	1541052
2	9.7	5557	140	75415	5015	766041
3	19	15053	920	569291	97	1380724
4	47	34371	590	355796	4301	654211
5	10	7510	210	122933	1150	599342

Such problems cannot be solved by special purpose knapsack algorithms (Pisinger)
(coefficients are too large).

Conclusions

- Generalization and tighter analysis of Aardal et. al reformulation technique.
- BR reformulation can be in range space, or null space – they seem equally effective on analyzed problems!
- Applicable to any IP (with equality/inequality constraints).
- Analysis of a general class of hard, infeasibleknapsack problems with decomposable coefficients.

Conclusions

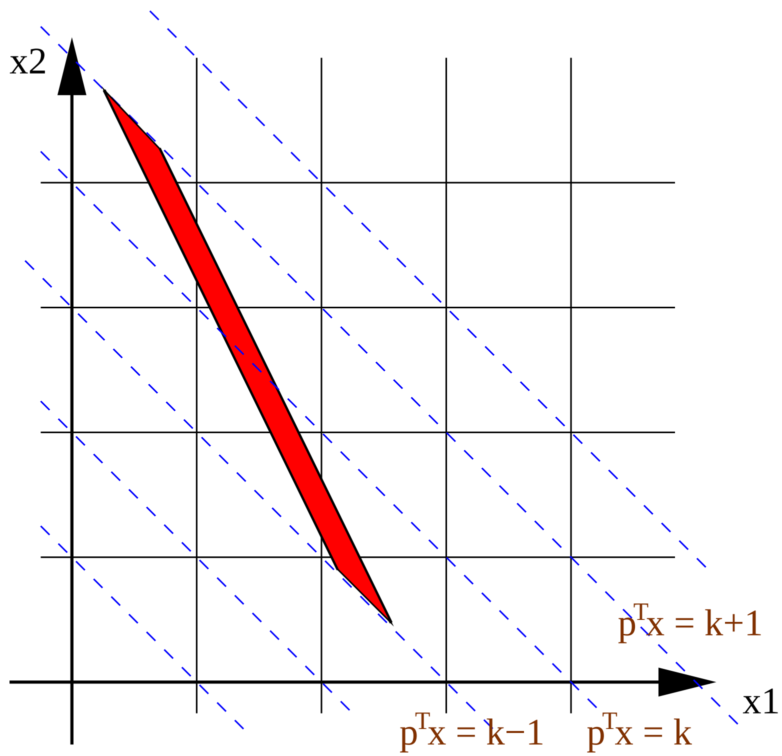
- Large values of big-M's: reason for *both* the provable hardness (in original space using x_js) and provable easiness (in reformulated space)!
- Choice of big-M's is too pessimistic as opposed to real life – both hardness and easiness holds for smaller values.
- Further work:
  - how about more than 1 dense constraint?
  - how does the method work on *any* IP?

Analysis of possibly feasible problems

$$(KP_2) \quad lb \leq a^T x \leq ub, \quad 0 \leq x \leq u, \quad x \in \mathcal{Z}^n$$

For some integer k, the feasible set of the LP relaxation is sandwiched between

$$p^T x = k - 1 \;\; and \;\; p^T x = k + 1.$$



Analysis of possibly feasible problems

- For a provably large number of B&B nodes, $\exists$ fractional solutions x_{lb} and x_{ub} with $a^T x_{lb} = lb$, $a^T x_{ub} = ub$.
- Branching on $p^T x$ tightens these bounds in 1 step; also y_n of the reformulation is $= p^T x$ in original problem.
- $p^T x$ may not be a "thin" direction! Possibly $0 \leq x_i \leq 1 \, \forall i$, and

$$\begin{aligned} \max \{ p^T x \,|\, x \text{ is LP-feasible} \} &- \min \{ p^T x \,|\, x \text{ is LP-feasible} \} \\ &= 1.9999, \end{aligned}$$

with all data integer; still
$p^T x$ is the **right** direction to branch on.

Example: Avis-Nemhauser problem

$$\text{Feasibility of} \quad \delta \leq a^T x \leq \gamma, \quad 0 \leq x \leq e, \; x \in \mathcal{Z}^n;$$
$$\text{with} \;\; a_i = n(n+1) + i$$

- γ, δ can be chosen such that problem is integer-infeasible and
- B&B takes a tree of depth at least $\frac{n-1}{2}$ to prove infeasibility.
- Branching on $e^T x$ kills the problem in one step
- Decomposing $a = pM + r$ gives $M = n(n+1)$, $r = (1 \; 2 \; \ldots \; n)$ and $p = e$!

Strongly correlated knapsack problem

– from C,U,W,W (1998)

$$\begin{array}{rlcl} \max & \sum_{i=0}^{5} \sum_{j=1}^{15} (100a_i + j - 1)\, x_{15i+j} & & \\ st. & \sum_{i=0}^{5} \sum_{j=0}^{14} a_i x_{15i+j} & \leq & \beta \\ & x_1, \ldots, x_{90} \in \{0, 1\} & & \end{array}$$

with

$$a^T = [\ 2222 \quad 1135 \quad 5555 \quad 2224 \quad 3334 \quad 1224\ ],\ \beta = 16999$$

Number of B&B nodes after column BR: 11.

Basis reduction in range space

- Find $n-1$ short, independent vectors in $L(A)$

$$g_2 = \begin{pmatrix} r_1p_2 - p_1r_2 \\ p_2 \\ -p_1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \ldots, g_n = \begin{pmatrix} r_1p_n - p_1r_n \\ p_n \\ 0 \\ 0 \\ \vdots \\ 0 \\ -p_1 \end{pmatrix}.$$

- short: $|g_i| \leq \max_{i=2,\ldots,n}\{\,|r_1p_i - r_ip_1| + |p_1| + |p_i|\,\}$

Number of B&B nodes in original formulation

Of independent interest!

- For the equality constrained knapsack $a^T x = a_0, \ x \geq 0$
- if known to be infeasible,

$$\# \text{ B\&B nodes } \geq \frac{4}{(n-1)!} \frac{a_0^{(n-1)}}{(a_2 a_3 \ldots a_n)}$$

- with $a = pM + r$, and $a_0 =$ (A & L lower bound on $F(a)$),

$$\# \text{ B\&B nodes } \geq C_1 \, (M + C_2)^{n-1}$$

C_1, C_2 constants indep. of M.