Parallel Approximation, and Integer Programming Reformulation

Gábor Pataki and Mustafa Tural *

Abstract

We analyze two integer programming reformulations of the knapsack feasibility problem

$$\begin{array}{rcl} \beta_1 & \leq ax \leq & \beta_2 \\ 0 & \leq x \leq & v \\ & x \in \mathbb{Z}^n & \end{array} \qquad (KP)$$

without assuming any structure on $a,$ only that its norm is large.

Both reformulations have a constraint matrix in which the columns form a reduced basis in the sense of Lenstra, Lenstra, and Lovász. The *nullspace reformulation* of Aardal, Hurkens and Lenstra has $n-1$ variables, and applies when $\beta_1 = \beta_2$. The *rangespace reformulation* of Krishnamoorthy and Pataki leaves the number of variables $n,$ and is applicable in general.

Assuming $\| a \| \geq 2^{(n/2+1)n}$ we prove an upper bound on the number of branch-and-bound nodes that are created, when branching on the last variable in the reformulations. The upper bound becomes 1, when $\| a \|$ is large enough.

The heart of the proof is an upper bound on the determinants of sublattices in LLL-reduced bases, and extracting a vector p which is "near parallel" to a from the transformation matrices. The near parallel vector is a good branching direction in (KP), and this transfers to the last variable in the reformulations.

Contents

*Department of Statistics and Operations Research, UNC Chapel Hill, **gabor@unc.edu, tural@email.unc.edu**

1 Introduction and notation

Geometry of Numbers and Integer Programming Starting with the work of H. W. Lenstra [15], algorithms based on the geometry of numbers have been an essential part of the Integer Programming landscape. Typically, these algorithms reduce an IP feasibility problem to a provably small number of smaller dimensional ones, and have strong theoretical properties. For instance, the algorithms of [15, 9, 16] have polynomial running time in fixed dimension; the algorithm of [6] has linear running time in dimension two. One essential tool in creating the subproblems is a "thin" branching direction, i.e. a c integral (row-)vector with the difference between the maximum and the minimum of cx over the underlying polyhedron being provably small. Basis reduction in lattices – in the Lenstra, Lenstra, Lovász (LLL) [14], or Korkine and Zolotarev (KZ) [10, 9] sense – is usually a key ingredient in the search for a thin direction. For implementations, and computational results, we refer to [4, 7, 18].

A simpler, and experimentally very successful technique for integer programming based on LLL-reduction was proposed by Aardal, Hurkens and A. K. Lenstra in [2] for equality constrained IP problems; see also [1]. Consider the problem

$$\begin{array}{rcl} Ax & = & b \\ 0 \leq & x & \leq v \\ x & \in & \mathbb{Z}^n, \end{array} \qquad \text{(IP-EQ)}$$

where A is an integral matrix with m independent rows, and let

$$\mathbb{N}(A) = \{ x \in \mathbb{Z}^n \mid Ax = 0 \}. \qquad (1.1)$$

The full-dimensional reformulation proposed in [2] is

$$\begin{array}{rcl} -x_b \leq & V\lambda & \leq v - x_b \\ \lambda & \in & \mathbb{Z}^{n-m}. \end{array} \qquad \text{(IP-EQ-N)}$$

Here V and x_b satisfy

$$\{ V\lambda \,|\, \lambda \in \mathbb{Z}^{n-m} \} \; = \; \mathbb{N}(A), \; x_b \in \mathbb{Z}^n, \; Ax_b = b,$$

the columns of V are reduced in the LLL-sense, and x_b is also short. For several classes of hard equality constrained integer programming problems – e.g. [5] – the reformulation turned out to be much easier to solve by commercial solvers than the original problem.

In [11] an even simpler, and experimentally just as effective reformulation method was introduced. It replaces

$$\begin{array}{rcl} b' \leq & Ax & \leq b \\ & x \in \mathbb{Z}^n & \end{array} \tag{IP}$$

with

$$\begin{array}{rcl} b' \leq & (AU)y & \leq b \\ & y \in \mathbb{Z}^n, & \end{array} \tag{IP-R}$$

where U is a unimodular matrix that makes the columns of AU reduced in the LLL-, or KZ-sense. It applies the same way, even if some of the inequalities in the IP feasibility problem are actually equalities. In [11] the authors also introduced a simplified method to compute a reformulation which is essentially equivalent to (IP-EQ-N).

We call (IP-R) the *rangespace reformulation* of (IP); and (IP-EQ-N) the *nullspace reformulation* of (IP-EQ).

These reformulation methods are very easy to describe (as opposed to say H. W. Lenstra's method), but seem difficult to analyze. The only analyses are for knapsack problems, with the weight vector having a given "decomposable" structure, i.e. $a = \lambda p + r$, with p, r, and λ integral, and λ large with respect to $\| p \|$, and $\| r \|$ – see [3, 11, 12].

The goal of this paper is to analyze these reformulations on the knapsack feasibility problem

$$\begin{array}{rcl} \beta_1 & \leq ax \leq & \beta_2 \\ 0 & \leq x \leq & v \\ & x \in \mathbb{Z}^n, & \end{array} \tag{KP}$$

where a is a positive, integral row vector, β_1, and β_2 are integers, without assuming any structure on the constraint vector *a priori*. We will assume only that $\| a \|$ is large – in fact, a key point will be that the large norm *implies* a decomposable structure, and this structure is automatically "discovered" by the reformulations.

The rangespace reformulation of (KP) is

$$\begin{array}{rcl} \beta_1 & \leq aUy \leq & \beta_2 \\ 0 & \leq Uy \leq & v \\ & y \in \mathbb{Z}^n, & \end{array} \tag{KP-R}$$

where U is a unimodular matrix that makes the columns of $\begin{pmatrix} a \\ I \end{pmatrix} U$ reduced in the LLL-sense (we do not analyze it with KZ-reduction). The nullspace reformulation is

$$\begin{array}{rcl} -x_\beta \leq & V\lambda & \leq v - x_\beta \\ \lambda & \in & \mathbb{Z}^{n-m}, \end{array} \qquad \text{(KP-N)}$$

where $x_\beta \in \mathbb{Z}^n$, $ax_\beta = \beta$, $\{ V\lambda \,|\, \lambda \in \mathbb{Z}^{n-m} \} = \mathbb{N}(a)$, and the columns of V are reduced in the LLL-sense.

Notation Vectors are column vectors, unless said otherwise. The ith unit row-vector is e_i. In general, when writing p_1, p_2, etc, we refer to vectors in a family of vectors. When p_i refers to the ith component of vector p, we will say this explicitly. For a rational vector b we denote by round(b) the vector obtained by rounding the components of b.

We will assume $0 \leq \beta_1 \leq \beta_2 \leq av$, and that the gcd of the components of a is 1.

For a polyhedron Q, and an integral row-vector c, the width, and the integer width of Q along c are

$$\begin{array}{rcl} \text{width}(c, Q) & = & \max\{ cx \,|\, x \in Q \} - \min\{ cx \,|\, x \in Q \}, \text{ and} \\ \text{iwidth}(c, Q) & = & \lfloor \max\{ cx \,|\, x \in Q \} \rfloor - \lceil \min\{ cx \,|\, x \in Q \} \rceil + 1. \end{array}$$

The integer width is the number of nodes generated by branch-and-bound when branching on the hyperplane cx; in particular, iwidth(e_i, Q) is the number of nodes generated when branching on x_i. If the integer width along any integral vector is zero, then Q has no integral points. Given an integer program labeled by (P), and c an integral vector, we also write width$(c, \text{(P)})$, and iwidth$(c, \text{(P)})$ for the width, and the integer width of the LP-relaxation of (P) along c, respectively.

A lattice in $\mathbb{R}^n$ is a set of the form

$$L = \mathbb{L}(B) = \{ Bx \,|\, x \in \mathbb{Z}^n \}, \qquad (1.2)$$

where B is a real matrix with n independent columns, called a *basis* of L. A square, integral matrix U is *unimodular* if $\det U = \pm 1$. It is well known that if B_1 and B_2 are bases of the same lattice, then $B_2 = B_1 U$ for some unimodular U. The determinant of L is

$$\det L = (\det B^T B)^{1/2}, \qquad (1.3)$$

where B is a basis of L; it is easy to see that $\det L$ is well-defined.

The LLL basis reduction algorithm [14] computes a reduced basis of a lattice in which the columns are "short" and "nearly" orthogonal. It runs in polynomial time for rational lattices. For simplicity, we use Schrijver's definition from [19]. Suppose that B has n independent columns, i.e.

$$B = [b_1, \ldots, b_n], \qquad (1.4)$$

and $b_1^*, \dots, b_n^*$ form the Gram-Schmidt orthogonalization of $b_1, \dots, b_n,$ that is $b_1 = b_1^*,$ and

$$b_i = b_i^* + \sum_{j=1}^{i-1} \mu_{ij} b_j^* \text{ with } \mu_{ij} = b_i^T b_j^* / \| b_j^* \|^2 \quad (i = 2, \dots, n;\ j \le i-1). \tag{1.5}$$

We call $b_1, \dots, b_n$ an *LLL-reduced basis of* $\mathbb{L}(B),$ if

$$|\mu_{ij}| \le 1/2 \quad (i = 2, \dots, n;\ j = 1, \dots, i-1), \text{ and} \tag{1.6}$$

$$\| b_i^* \|^2 \le 2 \| b_{i+1}^* \|^2 \quad (i = 1, \dots, n-1). \tag{1.7}$$

For an integral lattice $L,$ its *orthogonal lattice* is defined as

$$L^\perp = \{ y \in \mathbb{Z}^n \mid y^T x = 0\ \forall x \in L \},$$

and it holds that (see e.g. [17])

$$\det L^\perp \le \det L. \tag{1.8}$$

Suppose A is an integral matrix with independent rows. Then recalling (1.1), $\mathbb{N}(A)$ is the same as $\mathbb{L}(A^T)^\perp$. A lattice $L \subseteq \mathbb{Z}^n$ is called *complete*, if

$$L = \operatorname{lin} L \cap \mathbb{Z}^n.$$

The following lemma summarizes some basic results in lattice theory that we will use later on; for a proof, see for instance [17].

Lemma 1. *Let V be an integral matrix with n rows, and k independent columns, and $L = \mathbb{L}(V)$. Then (1) through (3) below are equivalent.*

(1) L is complete;

(2) $\det L^\perp = \det L$;

(3) There is a unimodular matrix Z s.t.

$$ZV = \begin{pmatrix} I_k \\ 0_{(n-k)\times k} \end{pmatrix}.$$

Furthermore, if Z is as in part (3), then the last $n-k$ rows of Z are a basis of $L^\perp$.

□

For an n-vector $a,$ we will write

$$\begin{aligned} f(a) &= 2^{n/4} / \| a \|^{1/n} \\ g(a) &= 2^{(n-2)/4} / \| a \|^{1/(n-1)}. \end{aligned} \tag{1.9}$$

2 Main results

In this section we will review the main results of the paper, give some examples, explanations, and some proofs that show their connection. The bulk of the work is the proof of Theorems 2, 3, 4, and 5, which is done in Section 3.

The main purpose of this paper is an analysis of the reformulation methods. This is done in Theorem 1, which proves an upper bound on the number of branch-and-bound nodes, when branching on the last variable in the reformulations. However, some of the intermediate results may be of interest on their own right.

In particular, Theorem 2 gives a bound on the determinant of a sublattice in an LLL-reduced basis, thus generalizing the well-known result from [14] showing that the first vector in such a basis is short.

Theorems 3 and 4 show that an integral vector p, which is "near parallel" to a can be extracted from the transformation matrices of the reformulations. The notion of near parallelness that we use is stronger than just requiring $\sin(a, p)$ to be small. The relationship of the two parallelness concepts is clarified in Proposition 1, and we will explain the connection with diophantine approximation in subsection 4.1.

Theorem 5 proves an upper bound on $\text{iwidth}(p, (KP))$, where p is an integral vector. A novelty of the bound is that it does not depend on β_1, and β_2, only on their difference. We show through examples that this bound is quite useful when p is a near parallel vector found according to Theorems 3 and 4.

In the end, a transference result between branching directions in the original, and reformulated problems completes the proof of Theorem 1.

Theorem 1. *Suppose* $\| a \| \geq 2^{(n/2+1)n}$. *Then*

(1) $\text{iwidth}(e_n, \text{ (KP-R)}) \leq \lfloor f(a)(2 \| v \| + (\beta_2 - \beta_1)) \rfloor + 1.$

(2) $\text{iwidth}(e_{n-1}, \text{ (KP-N)}) \leq \lfloor 2g(a) \| v \| \rfloor + 1.$

Our results focus on the *integer width*, not the *width*. The two differ by at most one, so when they are large, one can be used in place of the other. For instance, the algorithms of [15, 16] find a branching direction in which the width is bounded by an exponential function of the dimension. The goal is proving polynomial running time in fixed dimension, and this would still be achieved if the width were larger by a constant.

In contrast, when $\| a \|$ is sufficiently large, Theorem 1 implies that the integer width is at most *one* in both reformulations. It appears that such a result cannot be proven using the width, since

$$\text{width}(e_n, \text{ (KP-R)}) < 1 \tag{2.10}$$

does not seem to follow from any lower bound on $\| a \|$.

Theorem 2. *Suppose that $b_1, \ldots, b_n$ form an LLL-reduced basis of the lattice L, and denote by L_ℓ the lattice generated by $b_1, \ldots, b_\ell$. Then*

$$\det L_\ell \le 2^{\ell(n-\ell)/4} (\det L)^{\ell/n}. \tag{2.11}$$

Theorem 2 is a natural generalization of $\| b_1 \| \le 2^{(n-1)/4} (\det L)^{1/n}$ (see [14]). Despite its simplicity, Theorem 2 seems to be new.

The next two theorems show how a "near parallel" vector can be extracted from the transformation matrices used to find the reformulations.

Theorem 3. *Suppose $\| a \| \ge 2^{(n/2+1)n}$. Let U be a unimodular matrix such that the columns of*

$$\begin{pmatrix} a \\ I \end{pmatrix} U$$

are LLL-reduced, and p the last row of U^{-1}. Assume $\langle a, p \rangle > 0$; otherwise replace p by $-p$.

Let λp be the projection of a onto the line spanned by p, $r = a - \lambda p$. Then

(1) $\| p \| \, (1 + \| r \|^2)^{1/2} \le \| a \| \, f(a)$;

(2) $\lambda \ge 1/f(a)$;

(3) $\sin(a, p) \le \| r \| / \lambda \le 2f(a)$.

□

Theorem 4. *Suppose $\| a \| \ge 2^{(n/2+1)n}$. Let V be a matrix whose columns are an LLL-reduced basis of $\mathbb{N}(a)$, b a column vector with $ab = 1$, and p the $(n-1)st$ row of $(V, b)^{-1}$. Assume $\langle a, p \rangle > 0$; otherwise replace p by $-p$.*

Let λp be the projection of a onto the line spanned by p, $r = a - \lambda p$. Then $r \ne 0$, and

(1) $\| p \| \| r \| \le \| a \| \, g(a)$;

(2) $\sin(a, p) \le \| r \| / \lambda \le 2g(a)$.

□

It is important to note that p is integral, but λ and r may not be. Also, the measure of parallelness to a, i.e. the upper bound on $\| r \| / \lambda$ is quite similar for the p vectors found in Theorems 3 and 4, but their length can be quite different. When $\| a \|$ is large, the p vector in

Theorem 3 is guaranteed to be much shorter than a by $\lambda \geq 1/f(a)$. On the other hand, the p vector from Theorem 4 may be much *longer* than a : the upper bound on $\| p \| \| r \|$ does not guarantee any bound on $\| p \|$, since r can be fractional.

The following example illustrates this:

Example 1. *Consider the vector*

$$a = \begin{pmatrix} 3488, & 451, & 1231, & 6415, & 2191 \end{pmatrix}. \tag{2.12}$$

We computed p_1, r_1, λ_1 according to Theorem 3, and p_2, r_2, λ_2 according to Theorem 4.

$$\begin{aligned} p_1 &= \begin{pmatrix} 62, & 8, & 22, & 114, & 39 \end{pmatrix} \\ r_1 &= \begin{pmatrix} 0.2582, & 0.9688, & -6.5858, & 2.0554, & -2.9021 \end{pmatrix} \\ \lambda_1 &= 56.2539 \end{aligned} \tag{2.13}$$

$$\begin{aligned} p_2 &= \begin{pmatrix} 12204, & 1578, & 4307, & 22445, & 7666 \end{pmatrix} \\ r_2 &= \begin{pmatrix} -0.0165, & -0.0071, & 0.0194, & 0.0105, & -0.0140 \end{pmatrix} \\ \lambda_2 &= 0.2858 \end{aligned} \tag{2.14}$$

Clearly, p_1 is much shorter than a, while p_2 is much longer *than a. The measure of their parallelness to a is however, quite similar:*

$$\| r_1 \| / \lambda_1 = 0.1342, \; \textit{and} \; \| r_2 \| / \lambda_2 = 0.1110. \tag{2.15}$$

The following proposition clarifies the connection between two measures of parallelness, namely between $\| r \| / \lambda$, and $\sin(a, p)$ being small, and shows two further useful consequences of $\| r \| / \lambda$ being small.

Proposition 1. *Suppose that $a, p \in \mathbb{Z}^n$, λp is the projection of a onto the line spanned by p, and $r = a - \lambda p$. Assume $\lambda > 0$. Then the following hold:*

(1) $\sin(a, p) \leq \| r \| / \lambda$.

(2) For any M there is a, p with $\| a \| \geq M$ such that the inequality is strict.

(3) Denote by p_i and a_i the ith component of p, and a. If $\| r \| / \lambda < 1$, and $p_i \neq 0$, then the signs of p_i and a_i agree.

(4) If $\| r \| / \lambda < 1/2$, then $\mathrm{round}((1/\lambda) a) = p$.

Proof Statement (1) follows from

$$\sin(a, p) = \frac{\| r \|}{\| a \|} \leq \frac{\| r \|}{\| \lambda p \|} \leq \frac{\| r \|}{\lambda}, \tag{2.16}$$

where in the last inequality we used the integrality of p.

To see (2), consider the family of $a,$ and p vectors

$$\begin{aligned} a &= \left(m^2+1, \quad m^2\right), \\ p &= \left(m+1, \quad m\right) \end{aligned} \tag{2.17}$$

with m an integer. Letting λ and r be defined as in the statement of the proposition, a straightforward computation (or experimentation) shows that as $m \to \infty$

$$\begin{aligned} \sin(a,p) &\rightarrow 0, \\ \|r\| / \lambda &\rightarrow 1/\sqrt{2}. \end{aligned}$$

Statement (3) follows from

$$a_i = \lambda p_i + r_i = \lambda p_i \left(1 + \frac{r_i}{\lambda p_i}\right), \tag{2.18}$$

and

$$\left| \frac{r_i}{\lambda p_i} \right| \leq \frac{\|r\|}{\lambda}. \tag{2.19}$$

Finally, statement (4) follows from

$$(1/\lambda)a = p + (1/\lambda)r.$$

□

Theorem 5 below gives an upper bound on the number of branch-and-bound nodes when branching on a hyperplane in (KP).

Theorem 5. *Suppose that* $a = \lambda p + r,$ *with* $p \geq 0.$ *Then*

$$\text{iwidth}(p, (KP)) \leq \left\lfloor \frac{\|r\|\|v\|}{\lambda} + \frac{\beta_2 - \beta_1}{\lambda} \right\rfloor + 1. \tag{2.20}$$

This bound is quite strong for near parallel vectors computed from Theorems 3 and 4. For instance, let $a,$ $p_1,$ r_1, λ_1 be as in Example 1. If $\beta_1 = \beta_2$ in a knapsack problem with weight vector a, and each x_i is bounded between 0 and 11, then Theorem 5 implies that the integer width is at most one. At the other extreme, it also implies that the integer width is at most one, if each x_i is bounded between 0 and 1, and $\beta_2 - \beta_1 \leq 39.$ However, this bound does not seem as useful, when p is a "simple" vector, say a unit vector.

We now complete the proof of Theorem 1, based on a simple transference result between branching directions, taken from [11].

Proof of Theorem 1

Let us denote by Q, $\tilde{Q}$, and $\hat{Q}$ the feasible sets of the LP-relaxations of (KP), of (KP-R), and of (KP-N), respectively.

First, let U, and p be the transformation matrix, and the near parallel vector from Theorem 3. It was shown in [11] that $\text{iwidth}(p, Q) = \text{iwidth}(pU, \tilde{Q})$. But $pU = \pm e_n$, so

$$\text{iwidth}(p, Q) = \text{iwidth}(e_n, \tilde{Q}). \tag{2.21}$$

On the other hand,

$$\begin{aligned} \text{iwidth}(p, Q) &\leq \left\lfloor \frac{\| r \| \| v \|}{\lambda} + \frac{\beta_2 - \beta_1}{\lambda} \right\rfloor + 1 \\ &\leq \lfloor f(a)(2 \| v \| + (\beta_2 - \beta_1)) \rfloor + 1 \end{aligned} \tag{2.22}$$

with the first inequality coming from Theorem 5, and the second from using the bounds on $1/\lambda$ and $\| r \| / \lambda$ from Theorem 3. Combining (2.21) and (2.22) yields (1) in Theorem 1.

Now, let V, and p be the transformation matrix, and the near parallel vector from Theorem 4. It was shown in [11] that $\text{iwidth}(p, Q) = \text{iwidth}(pV, \hat{Q})$. But $pV = \pm e_{n-1}$, so

$$\text{iwidth}(e_{n-1}, \hat{Q}) = \text{iwidth}(p, Q). \tag{2.23}$$

On the other hand,

$$\begin{aligned} \text{iwidth}(p, Q) &\leq \left\lfloor \frac{\| r \| \| v \|}{\lambda} \right\rfloor + 1 \\ &\leq \lfloor g(a)(2 \| v \|) \rfloor + 1. \end{aligned} \tag{2.24}$$

with the first inequality coming from Theorem 5, and the second from using the bound on $\| r \| / \lambda$ in Theorem 4. Combining (2.23) and (2.24) yields (2) in Theorem 1.

3 Proofs

3.1 Sublattice determinants: proof of Theorem 2

Let $b_1^*, \ldots, b_n^*$ the Gram-Schmidt orthogonalization of $b_1, \ldots, b_n$, and write

$$\begin{aligned} \beta_i &:= \| b_i^* \|^2 \ (i = 1, \ldots, n), \\ D_\ell &:= (\det L_\ell)^2 = \beta_1 \cdot \dots \cdot \beta_\ell. \end{aligned} \tag{3.25}$$

The proof is by induction. For $\ell = n - 1$, multiplying the inequalities

$$\begin{aligned} \beta_n &\leq 2^0 \beta_n \\ \beta_{n-1} &\leq 2^1 \beta_n \\ &\vdots \\ \beta_1 &\leq 2^{n-1} \beta_n \end{aligned}$$

gives

$$D_n \le 2^{n(n-1)/2}\beta_n^{2n} \tag{3.26}$$

$$= 2^{n(n-1)/2}\left(\frac{D_n}{D_{n-1}}\right)^n \tag{3.27}$$

which after simplifying, yields

$$D_{n-1} \le 2^{(n-1)/2}(D_n)^{1-1/n}, \tag{3.28}$$

which is equivalent to the required result for $\ell = n-1$.

Suppose that (2.11) is true for $\ell \le n-1$; we will prove it for $\ell - 1$. Since $b_1, \ldots, b_\ell$ forms an LLL-reduced basis of L_ℓ, we can replace n by ℓ in (3.28) to get

$$D_{\ell-1} \le 2^{(\ell-1)/2}(D_\ell)^{(\ell-1)/\ell}. \tag{3.29}$$

By the induction hypothesis,

$$D_\ell \le 2^{\ell(n-\ell)/2}(D_n)^{\ell/n}, \tag{3.30}$$

from which we obtain

$$(D_\ell)^{(\ell-1)/\ell} \le 2^{(\ell-1)(n-\ell)/2}(D_n)^{(\ell-1)/n}. \tag{3.31}$$

Using the upper bound on $(D_\ell)^{(\ell-1)/\ell}$ from (3.31) in (3.29) yields

$$D_{\ell-1} \le 2^{(\ell-1)/2}2^{(\ell-1)(n-\ell)/2}(D_n)^{(\ell-1)/\ell} \tag{3.32}$$

$$= 2^{(\ell-1)(n-(\ell-1))/2}(D_n)^{(\ell-1)/n}, \tag{3.33}$$

as required.

□

3.2 Near parallel vectors: intuition, and proofs for Theorems 3 and 4

Intuition for Theorem 3 We review a proof from [11], which applies when we know *a priori* the existence of a decomposition

$$a = p\lambda + r, \tag{3.34}$$

with λ large with respect to $\|p\|$, and $\|r\|$. The reason that the columns of

$$\begin{pmatrix} a \\ I \end{pmatrix} = \begin{pmatrix} \lambda p + r \\ I \end{pmatrix}$$

are *not* short and orthogonal is the presence of the $\lambda_i p_i$ components in the first row. So if postmultiplying by a unimodular U results in reducedness, it is natural to expect that many components of pU will be zero; indeed it follows from the properties of LLL-reduction, that the first $n-1$ components *will* be zero. Since U has full rank, the nth component of pU must be nonzero. So p will be the a multiple of the last row of U^{-1}, in other words, the last row of U^{-1} will be near parallel to a. (In [11] it was assumed that p, r, and λ are integral, but the proof would work even if λ and r were rational.)

It is then natural to expect that the last row of U^{-1} will give a near parallel vector to a, even if a decomposition like (3.34) is not known in advance. This is indeed what is shown in Theorem 3, when $\| a \|$ is sufficiently large.

Proof of Theorem 3 First note that the lower bound on $\| a \|$ implies

$$f(a) \leq \sqrt{3}/2. \tag{3.35}$$

Let L_ℓ be the lattice generated by the first ℓ columns of $\begin{pmatrix} a \\ I \end{pmatrix} U$, and

$$Z = \begin{pmatrix} 0 & U^{-1} \\ 1 & -a \end{pmatrix}.$$

Clearly, Z is unimodular, and

$$Z \begin{pmatrix} aU \\ U \end{pmatrix} = \begin{pmatrix} I_n \\ 0_{1\times n} \end{pmatrix}. \tag{3.36}$$

So Lemma 1 implies that L_ℓ is complete, and the last $n+1-\ell$ rows of Z generate $L_\ell^\perp$. The last row of Z is $(1,-a)$, and the next-to-last is $(0,p)$, so we get

$$\begin{aligned} \det L_n &= \det L_n^\perp = (\| a \|^2 + 1)^{1/2}, \\ \det L_{n-1} &= \det L_{n-1}^\perp = \| p \| \, (1 + \| r \|^2)^{1/2}. \end{aligned} \tag{3.37}$$

Theorem 2 implies

$$\det L_{n-1} \leq 2^{(n-1)/4} (\det L_n)^{1-1/n}. \tag{3.38}$$

Substituting into (3.38) from (3.37) gives

$$\begin{aligned} \| p \| \, (1 + \| r \|^2)^{1/2} &\leq 2^{(n-1)/4} (\sqrt{\| a \|^2 + 1})^{1-1/n} \\ &\leq 2^{n/4} \, \| a \|^{1-1/n} \\ &= \| a \| \, f(a), \end{aligned} \tag{3.39}$$

with the second inequality coming the lower bound on $\| a \|$. This shows (1).

Proof of (2) From (1) we directly obtain

$$\begin{aligned}\frac{f(a)^2 \, \|a\|^2 - \|r\|^2}{\|p\|^2} &\geq \frac{f(a)^2 \, \|a\|^2 - \|p\|^2 \|r\|^2}{\|p\|^2} \\ &\geq 1 \\ &= \frac{f(a)^2 \, \|a\|^2}{f(a)^2 \, \|a\|^2},\end{aligned} \tag{3.40}$$

where in the first inequality we used $\|p\| \geq 1$. Now note

$$\|p\|^2 \leq f(a)^2 \, \|a\|^2,$$

i.e. the the denominator of the first expression in (3.40) is not larger than the denominator of the last expression. So if we replace $f(a)^2$ by 1 in the *numerator* of both, the inequality will remain valid. The result is

$$\frac{\|a\|^2 - \|r\|^2}{\|p\|^2} \geq \frac{1}{f(a)^2}, \tag{3.41}$$

which is the square of the required inequality.

Proof of (3) We have

$$\begin{aligned}\sin(a,p)^2 &\leq \frac{\|r\|^2}{\lambda^2} \\ &= \frac{\|p\|^2 \|r\|^2}{\|\lambda p\|^2} \\ &\leq \frac{\|p\|^2 \|r\|^2}{\|a\|^2 - \|r\|^2} \\ &\leq \frac{f(a)^2 \, \|a\|^2}{\|a\|^2 - \|r\|^2} \\ &\leq \frac{f(a)^2 \, \|a\|^2}{\|a\|^2 - f(a)^2 \, \|a\|^2} \\ &= \frac{f(a)^2}{1 - f(a)^2} \\ &\leq 4f(a)^2,\end{aligned} \tag{3.42}$$

where the first inequality comes from Proposition 1, the last from (3.35), and the others are straightforward;

□

Intuition for Theorem 4 We recall a proof from [11], which applies when we know *a priori* the existence of a decomposition like in (3.34) with λ large with respect to $\|p\|$, and $\|r\|$, and p not a multiple of r. It is shown there that the first $n-2$ components of pV will be zero. Denote by L_ℓ the lattice generated by the first ℓ columns of V. So p is in $L_{n-2}^{\perp}$, and it is not a multiple of a, but it is near parallel to it.

So one can expect that an element of $L_{n-2}^{\perp}$ which is distinct from a will be near parallel to a, even if a decomposition like (3.34) is not known in advance. The p described in Theorem 4 will be such a vector.

Proof of Theorem 4 The lower bound on $\| a \|$ implies

$$g(a) \le \sqrt{3}/2. \tag{3.43}$$

As noted above, let L_ℓ be the lattice generated by the first ℓ columns of V. We have

$$(V,b)^{-1} V \;=\; \begin{pmatrix} I_{n-1} \\ 0 \end{pmatrix}. \tag{3.44}$$

So Lemma 1 implies that L_ℓ is complete, and the last $n-\ell$ rows of $(V,b)^{-1}$ generate $L_\ell^{\perp}$. It is elementary to see that the last row of $(V,b)^{-1}$ is a, and by definition the next-to-last row is p, and these rows are independent, so $r \neq 0$. Also,

$$\begin{aligned} \det L_{n-1} &= \det L_{n-1}^{\perp} = \| a \|, \\ \det L_{n-2} &= \det L_{n-2}^{\perp} = \| p \| \| r \| . \end{aligned} \tag{3.45}$$

Theorem 2 with $n-1$ in place of n, and $n-2$ in place of ℓ implies

$$\det L_{n-2} \;\le\; 2^{(n-2)/4} (\det L_{n-1})^{1-1/(n-1)}. \tag{3.46}$$

Substituting into (3.46) from (3.45) gives

$$\begin{aligned} \| p \| \| r \| &\le 2^{(n-2)/4} \; \| a \|^{1-1/(n-1)} \\ &= \| a \| \; g(a), \end{aligned} \tag{3.47}$$

as required.

Proof of (2) It is enough to note that in proof of (3) in Theorem 3 we only used the inequality $\| p \|^2 \| r \|^2 \le f(a)^2 \; \| a \|^2$. So the exact same argument works here as well with $g(a)$ instead of $f(a)$, and invoking (3.43) as well.

□

3.3 Branching on a near parallel vector: proof of Theorem 5

This proof is somewhat technical, so we state, and prove some intermediate claims, to improve readability. Let us fix a, p, β_1, β_2, and v. For a row-vector w, and an integer ℓ we write

$$\begin{aligned} \max(w,\ell) &= \max \{ \, wx \mid px \le \ell, \, 0 \le x \le v \, \} \\ \min(w,\ell) &= \min \{ \, wx \mid px \ge \ell, \, 0 \le x \le v \, \}. \end{aligned} \tag{3.48}$$

The dependence on p, on v, and on the sense of the constraint (i.e. $\leq$, or $\geq$) is not shown by this notation; however, we always use $px \leq \ell$ with "max", and $px \geq \ell$ with "min", and p and v are fixed. Note that as a is a row-vector, and v a column-vector, av is their inner product, and the meaning of pv is similar.

Claim 1. *Suppose that ℓ_1 and ℓ_2 are integers in $\{0, \ldots, pv\}$. Then*

$$\min(a, \ell_2) - \max(a, \ell_1) \quad \geq \quad - \| r \| \| v \| + \lambda(\ell_2 - \ell_1). \tag{3.49}$$

Proof The decomposition of a shows

$$\begin{aligned} \max(a, \ell_1) &\leq \max(r, \ell_1) + \lambda\ell_1, \text{ and} \\ \min(a, \ell_2) &\geq \min(r, \ell_2) + \lambda\ell_2. \end{aligned} \tag{3.50}$$

So we get the following chain of inequalities, with ensuing explanation:

$$\begin{aligned} \min(a, \ell_2) - \max(a, \ell_1) &\geq \min(r, \ell_2) - \max(r, \ell_1) + \lambda(\ell_2 - \ell_1) \\ &\geq rx_2 - rx_1 + \lambda(\ell_2 - \ell_1) \\ &= r(x_2 - x_1) + \lambda(\ell_2 - \ell_1) \\ &\geq - \| r \| \| v \| + \lambda(\ell_2 - \ell_1). \end{aligned} \tag{3.51}$$

Here x_2 and x_1 are the solutions that attain the maximum, and the minimum in $\min(r, \ell_2)$ and $\max(r, \ell_1)$, respectively. The last inequality follows from the fact that the ith component of $x_2 - x_1$ is at most v_i in absolute value, and the Cauchy-Schwartz inequality.

End of proof of Claim 1

Next, let us note

$$\min(a, k) \quad \leq \quad \max(a, k) \text{ for } k \in \{0, \ldots, pv\}. \tag{3.52}$$

Indeed, (3.52) holds, since the feasible sets of the optimization problems defining $\min(a, k)$, and $\max(a, k)$ contain $\{ x \,|\, px = k, \, 0 \leq x \leq v \}$.

The nonnegativity of p and of a imply $\min(a, 0) = 0$, and $\max(a, pe) = av$. The proof of the following claim is trivial, hence omitted.

Claim 2. *Suppose that ℓ_1 and ℓ_2 are integers in $\{0, \ldots, pv\}$ with $\ell_1 + 1 \leq \ell_2$, and*

$$\max(a, \ell_1) < \beta_1 \leq \beta_2 < \min(a, \ell_2). \tag{3.53}$$

Then for all x with $\beta_1 \leq ax \leq \beta_2$, $0 \leq x \leq v$

$$\ell_1 < px < \ell_2 \tag{3.54}$$

holds.

We assume for simplicity

$$\max(a, 0) < \beta_1 \leq \beta_2 < \min(a, pe); \tag{3.55}$$

the cases when this fails to hold are easy to handle separately. Let ℓ_1 be the largest, and ℓ_2 the smallest integer such that

$$\max(a, \ell_1) < \beta_1 \leq \beta_2 < \min(a, \ell_2). \tag{3.56}$$

From (3.52) $\ell_2 \geq \ell_1 + 1$ follows, and Claim 2 yields

$$\text{iwidth}(p,\ (\text{KP})) \ \leq\ \ell_2 - \ell_1 - 1. \tag{3.57}$$

By the choices of ℓ_1, and ℓ_2 we have

$$\beta_1 \leq \max(a, \ell_1 + 1),\ \text{and}\ \beta_2 \geq \min(a, \ell_2 - 1), \tag{3.58}$$

hence Claim 1 leads to

$$\begin{aligned} \beta_2 - \beta_1 &\geq \min(a, \ell_2 - 1) - \max(a, \ell_1 + 1) \\ &\geq -\| r \| \| v \| + \lambda(\ell_2 - \ell_1 - 2), \end{aligned} \tag{3.59}$$

that is

$$\ell_2 - \ell_1 - 2 \ \leq\ \frac{\beta_2 - \beta_1}{\lambda} + \frac{\| r \| \| v \|}{\lambda}. \tag{3.60}$$

Comparing (3.57) and (3.60) yields completes the proof.

□

4 Discussion

4.1 Connection with diophantine approximation, and other notions of near parallelness

Given a rational vector b, simultaneous diophantine approximation (see e.g. [14, 13]) computes an integral vector p, and an integer q, such that q, and $\| b - (1/q)p \|$ are both small. Suppose now that given an *integral* vector a, we apply diophantine approximation to $(1/\mu)a$, where μ is a rational, then set $\lambda = \mu/q$, $r = a - \lambda p$. Then $\| r \| / \lambda$ will be small, and if μ is suitably chosen, say $\mu = \| a \|$, then $\| p \|$ will be small as well.[1].

So computing a near parallel vector can be done in other ways as well. The relevance of Theorems 3 and 4 is not just finding near parallel vectors: it is finding a near parallel p, which corresponds to a unit vector in the rangespace- and nullspace reformulations, thus leading to the analysis of Theorem 1.

[1]Thanks are due to Laci Lovász and Fritz Eisenbrand for pointing out this connection

Finding an integral vector, which is near parallel to an other integral or rational one has other applications as well. In [8] Huyer, and Neumaier studied several notions of near parallelness, presented numerical algorithms, and applications to verifying the feasibility of a linear system of inequalities.

4.2 Successive approximation

Theorems 3 and 4 approximate a by a single vector. One needs to modify the proofs only slightly to obtain results in which a is approximated by a linear combination of integral vectors. As of now, we don't know how to use the general results for a better analysis of the reformulations than what is already given in Theorem 1. However, there is a natural geometric intuition behind them: if one row of U^{-1} or of $(V,b)^{-1}$ gives a good approximation of a, then a combination of $2,3,\ldots,k$ must give increasingly better approximations. Since Theorems 6 and 7 verify this intuition, it is worth stating them, and outlining the proofs.

Let us define

$$\begin{array}{rcl} f(a,k) & = & 2^{(k(n-k)+1)/4} / \parallel a \parallel^{k/n} \\ g(a,k) & = & 2^{k(n-1-k)/4} / \parallel a \parallel^{(k-1)/n} . \end{array} \tag{4.61}$$

The successive version of Theorem 3 is given below:

Theorem 6. *Let $a \in \mathbb{Z}^n$ be a row-vector, with $\parallel a \parallel \geq 2^{(n/2+1)n}$, U a unimodular matrix such that the columns of*

$$\begin{pmatrix} a \\ I \end{pmatrix} U$$

are LLL-reduced, and P_k the (integral) submatrix of U^{-1} consisting of the last k rows. Furthermore, let $a(k)$ be the projection of a onto the subspace spanned by the rows of P_k, $r = a - a(k)$, and

$$\lambda_k := \parallel a(k) \parallel / \det(P_k P_k^T)^{1/2}.$$

Then

(1) $(\det(P_k P_k^T))^{1/2}(1+ \parallel r \parallel^2)^{1/2} \leq \parallel a \parallel f(a,k)$;

(2) $\lambda_k \geq 1/f(a,k)$;

(3) $|\sin(a, a(k))| \leq \parallel r \parallel / \lambda_k \leq 2f(a,k)$.

Proof sketch We will use the notation of Theorem 3. In its proof we simply change (3.37) (we copy the first expression for $\det L_n$ for easy reference) to

$$\begin{array}{rcl} \det L_n & = & \det L_n^{\perp} = (\parallel a \parallel^2 + 1)^{1/2}, \\ \det L_{n-k} & = & \det L_{n-k}^{\perp} = (\det(P_k P_k^T))^{1/2}(1+ \parallel r \parallel^2)^{1/2}, \end{array} \tag{4.62}$$

and (3.38) to

$$\det\, L_{n-k} \;\; \leq \;\; 2^{k(n-k)/4} (\det L_n)^{1-k/n}. \tag{4.63}$$

Then substituting into (4.63) from (4.62) gives

$$\begin{aligned} (\det(P_k P_k^T))^{1/2} (1+ \|\, r \,\|^2)^{1/2} \;\; &\leq \;\; 2^{(k(n-k))/4} (\sqrt{\|\, a \,\|^2 +1})^{1-k/n} \\ &\leq \;\; 2^{(k(n-k)+1)/4} / \, \|\, a \,\|^{k/n} \\ &= \;\; \|\, a \,\| \; f(a,k), \end{aligned} \tag{4.64}$$

with the second inequality coming the lower bound on $\|\, a \,\|$. This shows (1), and the rest of the proof follows verbatim the proof of Theorem 3. □

Theorem 4 also has a successive variant, which is

Theorem 7. *Suppose* $\|\, a \,\| \geq \; 2^{(n/2+1)n}$. *Let* V *be a matrix whose columns are an LLL-reduced basis of* $\mathbb{N}(a)$, b *a column vector with* $ab = 1$, $k \leq n-1$ *an integer, and* P_k *the (integral) submatrix of* $(V, b)^{-1}$ *consisting of the last* k *rows.*

Furthermore, let $a(k)$ *be the projection of* a *onto the subspace spanned by the rows of* P_k, $r = a - a(k)$, *and*

$$\lambda_k := \|\, a(k) \,\| \; / \det(P_k P_k^T)^{1/2}.$$

Then $r \neq 0$, *and*

(1) $(\det(P_k P_k^T))^{1/2} \; \|\, r \,\| \leq \|\, a \,\| \; g(a,k)$;

(2) $|\sin(a, a(k))| \leq \|\, r \,\| \; / \lambda \leq 2g(a,k)$.

Proof sketch We will use the notation of Theorem 4. We need to replace (3.45) with

$$\begin{aligned} \det L_{n-1} \;\; &= \;\; \det L_{n-1}^{\perp} \; = \|\, a \,\|, \\ \det L_{n-1-k} \;\; &= \;\; \det L_{n-1-k}^{\perp} \; = \; (\det(P_k P_k^T))^{1/2} \; \|\, r \,\| \; . \end{aligned} \tag{4.65}$$

Theorem 2 with $n-1$ in place of n, and $n-1-k$ in place of ℓ implies

$$\det\, L_{n-1-k} \;\; \leq \;\; 2^{k(n-1-k)/4} (\det L_{n-1})^{1-k/(n-1)}. \tag{4.66}$$

Plugging the expressions for $\det L_{n-1}$ and $\det L_{n-1-k}$ from (4.65) into (4.66) gives

$$\begin{aligned} (\det(P_k P_k^T))^{1/2} \; \|\, r \,\| \;\; &\leq \;\; 2^{k(n-1-k)/4} \; \|\, a \,\|^{1-k/(n-1)} \\ &= \;\; g(a,k) \; \|\, a \,\|, \end{aligned} \tag{4.67}$$

proving (1). The rest of the proof is an almost verbatim copy of the corresponding proof in Theorem 4. □

Acknowledgement We thank Don Coppersmith for his generous, and kind help on the $n = 2$ case; Ravi Kannan for helpful discussions; and Laci Lovász and Fritz Eisenbrand for discussions on the connection with diophantine approximation.

References

[1] Karen Aardal, Robert E. Bixby, Cor A. J. Hurkens, Arjen K. Lenstra, and Job W. Smeltink. Market split and basis reduction: Towards a solution of the Cornuéjols-Dawande instances. *INFORMS Journal on Computing*, 12(3):192–202, 2000.

[2] Karen Aardal, Cor A. J. Hurkens, and Arjen K. Lenstra. Solving a system of linear Diophantine equations with lower and upper bounds on the variables. *Mathematics of Operations Research*, 25(3):427–442, 2000.

[3] Karen Aardal and Arjen K. Lenstra. Hard equality constrained integer knapsacks. *Mathematics of Operations Research*, 29(3):724–738, 2004.

[4] William Cook, Thomas Rutherford, Herbert E. Scarf, and David F. Shallcross. An implementation of the generalized basis reduction algorithm for integer programming. *ORSA Journal on Computing*, 5(2):206–212, 1993.

[5] Gérard Cornuéjols and Milind Dawande. A class of hard small 0–1 programs. In *6th Conference on Integer Programming and Combinatorial Optimization*, volume 1412 of *Lecture notes in Computer Science*, pages 284–293. Springer-Verlag, 1998.

[6] Friedrich Eisenbrand and Sören Laue. A linear algorithm for integer programming in the plane. *Mathematical Programming*, 102(2):249–259, 2005.

[7] Liyan Gao and Yin Zhang. Computational experience with lenstra's algorithm. *Technical Report, Department of Computational and Applied Mathematics, Rice University*, 2002.

[8] Walfred Huyer and Arnold Neumaier. Integral approximation of rays and verification of feasibility. *Reliable Computing*, 10:195–207, 2004.

[9] Ravi Kannan. Minkowski's convex body theorem and integer programming. *Mathematics of Operations Research*, 12(3):415–440, 1987.

[10] A. Korkine and G. Zolotarev. Sur les formes quadratiques. *Mathematische Annalen*, 6:366–389, 1873.

[11] Bala Krishnamoorthy and Gábor Pataki. Column basis reduction and decomposable knapsack problems. *Research Report, Dept of Statistics and Operations Research, UNC-Chapel Hill, submitted.*, 2006.

[12] Bala Krishnamoorthy and Gábor Pataki. Column basis reduction and higher level decomposable knapsack problems. *Research Report, Dept of Statistics and Operations Research, UNC-Chapel Hill, submitted.*, 2007.

[13] Jeffrey C. Lagarias. The computational complexity of simultaneous diophantine approximation. *SIAM J. Comput.*, 14:196–209, 1985.

[14] Arjen K. Lenstra, Hendrik W. Lenstra, Jr., and László Lovász. Factoring polynomials with rational coefficients. *Mathematische Annalen*, 261:515–534, 1982.

[15] Hendrik W. Lenstra, Jr. Integer programming with a fixed number of variables. *Mathematics of Operations Research*, 8:538–548, 1983.

[16] László Lovász and Herbert E. Scarf. The generalized basis reduction algorithm. *Mathematics of Operations Research*, 17:751–764, 1992.

[17] Jacques Martinet. *Perfect Lattices in Euclidean Spaces.* Springer-Verlag, Berlin, 2003.

[18] Sanjay Mehrotra and Zhifeng Li. On generalized branching methods for mixed integer programming. *Research Report, Department of Industrial Engineering, Northwestern University*, 2004.

[19] Alexander Schrijver. *Theory of Linear and Integer Programming.* Wiley, Chichester, United Kingdom, 1986.