On the Structure of Reduced Kernel Lattice Bases

Karen Aardal[1,2] and Frederik von Heymann[1]

[1] Delft Institute of Applied Mathematics, TU Delft, The Netherlands
{k.i.aardal,f.j.vonheymann}@tudelft.nl
[2] Centrum Wiskunde en Informatica, Amsterdam, The Netherlands

Abstract. Lattice-based reformulation techniques have been used successfully both theoretically and computationally. One such reformulation is obtained from the lattice $\ker_{\mathbb{Z}}(\boldsymbol{A}) = \{\boldsymbol{x} \in \mathbb{Z}^n \mid \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}\}$. Some of the hard instances in the literature that have been successfully tackled by lattice-based techniques, such as market split and certain classes of knapsack instances, have randomly generated input $\boldsymbol{A}$. These instances have been posed to stimulate algorithmic research. Since the considered instances are very hard even in low dimension, less experience is available for larger instances. Recently we have studied larger instances and observed that the LLL-reduced basis of $\ker_{\mathbb{Z}}(\boldsymbol{A})$ has a specific sparse structure. In particular, this translates into a map in which some of the original variables get a "rich" translation into a new variable space, whereas some variables are only substituted in the new space. If an original variable is important in the sense of branching or cutting planes, this variable should be translated in a non-trivial way. In this paper we partially explain the obtained structure of the LLL-reduced basis in the case that the input matrix $\boldsymbol{A}$ consists of one row $\boldsymbol{a}$. Since the input is randomly generated our analysis will be probabilistic. The key ingredient is a bound on the probability that the LLL algorithm will interchange two subsequent basis vectors. It is worth noticing that computational experiments indicate that the results of this analysis seem to apply in the same way also in the general case that $\boldsymbol{A}$ consists of multiple rows. Our analysis has yet to be extended to this general case. Along with our analysis we also present some computational indications that illustrate that the probabilistic analysis conforms well with the practical behavior.

1 Introduction

Consider the following integer program:

$$\max\{\boldsymbol{c}\boldsymbol{x} \mid \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},\ \boldsymbol{x} \geq \boldsymbol{0}\}, \tag{1}$$

where $\boldsymbol{A}$ is an integer $m \times n$ matrix of full row rank and $\boldsymbol{b}$ an integer m-vector. Starting with the well-known algorithm of Lenstra [13], several lattice-based approaches to reformulate the feasible region have been proposed, see, e.g., [1, 3, 5, 11, 16–18]. Here we will consider the reformulation as in [1]:

$$\boldsymbol{x} := \boldsymbol{x}^0 + \boldsymbol{Q}\boldsymbol{\lambda}\,, \tag{2}$$

where $\boldsymbol{x}^0 \in \mathbb{Z}^n$ satisfies $\boldsymbol{A}\boldsymbol{x}^0 = \boldsymbol{b}$, $\boldsymbol{\lambda} \in \mathbb{Z}^{n-m}$, and $\boldsymbol{Q}$ is a basis for the lattice $\ker_{\mathbb{Z}}(\boldsymbol{A}) = \{\boldsymbol{x} \in \mathbb{Z}^n \mid \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}\}$. Due to the nonnegativity requirements on the $\boldsymbol{x}$-variables, one now obtains an equivalent formulation of the integer program (1):

$$\max\{\boldsymbol{c}(\boldsymbol{x}^0 + \boldsymbol{Q}\boldsymbol{\lambda}) \mid Q\boldsymbol{\lambda} \geq -\boldsymbol{x}^0\}\,. \tag{3}$$

This reformulation has been shown to be of particular computational interest in the case where $\boldsymbol{Q}$ is reduced in the sense of Lovász [12].

Several authors have studied knapsack instances that have a particular structure that makes them particularly difficult to solve by "standard" methods such as branch-and-bound. Examples of such instances can be found in [2, 7, 11]. Common for these instances is that the input is generated in such a way that the resulting lattice $\ker_{\mathbb{Z}}(\boldsymbol{A})$ has a very particular structure that makes the reformulated instances almost trivial to solve. Other instances that are randomly generated without any particular structure of the $\boldsymbol{A}$-matrix, such as the market split instances [6] and knapsack instances studied in [2, 3], have no particular lattice structure. Yet they are practically unsolvable by branch-and-bound in the original $\boldsymbol{x}$-variable space, whereas their lattice reformulation solves rather easily, at least up to a certain dimension. It is still to be understood why the lattice reformulation for these instances is computationally more effective.

If we consider the randomly generated instance without any particular lattice structure and solve small instances, such as $n - m \leq 25$, one typically observes that the number of zeros in the basis $\boldsymbol{Q}$ is small. In higher dimension, and here "high" is depending on the input, a certain sparser structure will start to appear.

More specifically, we observe computationally that $\boldsymbol{Q}$ has a certain number of rows with rich interaction between the variables $\boldsymbol{x}$ and $\boldsymbol{\lambda}$, but from some point on this interaction breaks down almost instantly and we get one '1' per row, i.e., $\boldsymbol{Q}$ yields variable substitutions. To be able to better understand the relative effectiveness of the lattice reformulation, and in order to be able to apply the lattice reformulation in a (more) useful way in higher dimension, it is important to identify the variables that have a nontrivial translation into the new $\boldsymbol{\lambda}$-variable space.

In this paper we partially explain the phenomenon described above for the case that $m = 1$, that is, $\boldsymbol{A}$ consists of a single row $\boldsymbol{a} = (a_1, \ldots, a_n)$. As the exact structure of $\boldsymbol{Q}$ depends on the choice of $\boldsymbol{a}$, our analysis will be probabilistic. To this end, we assume that the entries of our input vector $\boldsymbol{a}$ are drawn independently and uniformly at random from an interval $[l, \ldots, u] := [l, u] \cap \mathbb{Z}$, where $0 < l < u$. We notice that explaining the phenomenon is related to the analysis of the probability that the LLL-algorithm performs a basis vector interchange after a basis vector with a certain index k has been considered by the algorithm.

Let $\boldsymbol{Q} = [\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{n-1}]$ be an LLL y-reduced basis (see Section 2 for more details) of $\ker_{\mathbb{Z}}(\boldsymbol{a})$, and let $\boldsymbol{b}_1^*, \ldots, \boldsymbol{b}_{n-1}^*$ be the Gram-Schmidt vectors corresponding to $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{n-1}$. If $\|\boldsymbol{b}_{i+1}^*\|^2 \geq y\|\boldsymbol{b}_i^*\|^2$, then basis vectors $i + 1$ and i will not be interchanged. We will show that, starting with a basis $\bar{\boldsymbol{Q}}$ of $\ker_{\mathbb{Z}}(\boldsymbol{a})$ of a certain structure, the probability that the LLL-algorithm [12] performs basis vector interchanges becomes increasingly small the higher the index of the basis vector. In particular, for given l, u, and reduction factor y, we derive a constant c and a k_0, such that for $k \geq k_0$ we have

$$\Pr\left(\|\boldsymbol{b}_{k+1}^*\|^2 < y\|\boldsymbol{b}_k^*\|^2\right) \quad \leq \quad e^{-c(k+1)^2} + 2^{-(k+1)/2}. \tag{4}$$

Note that stated in this form it is an asymptotic result, but we will see that the values of k_0 are very similar to the ones observed in the experiments.

To derive a bound on $\Pr\left(\|\boldsymbol{b}_{k+1}^*\|^2 < y\|\boldsymbol{b}_k^*\|^2\right)$ we first need to be able to express the length of the Gram-Schmidt vectors $\boldsymbol{b}_j^*$ in terms of the input vector $\boldsymbol{a}$. This is done in Section 2 and results in Expression (18). The bound on $\Pr\left(\|\boldsymbol{b}_{k+1}^*\|^2 < y\|\boldsymbol{b}_k^*\|^2\right)$ is derived through several steps in Section 3. In this derivation, the challenge is that $\|\boldsymbol{b}_{k+1}^*\|^2$ and $\|\boldsymbol{b}_k^*\|^2$ are not independent. To estimate the mean of the ratio $\|\boldsymbol{b}_{k+1}^*\|^2/\|\boldsymbol{b}_k^*\|^2$, we use a result by Pittenger [19], and to estimate how much this ratio deviates from the mean we use the Azuma-Hoeffding inequality [4, 8]. Some computational indications and further discussion are provided in Section 4. We notice that the computational results corresponds well to the observed practical behavior of the LLL algorithm on the considered class of input.

2 Notation and preliminaries

We first repeat some known facts about lattices and bases of lattices, as well as a high-level description of the LLL-algorithm. Then we give some properties of the kernel lattice of $\boldsymbol{a}$.

2.1 Basic results on lattices

Let L be a lattice in $\mathbb{R}^n$, i.e., a discrete additive subgroup of $\mathbb{R}^n$. Furthermore, let $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_m,\ m \leq n$, be a basis of L, and let $\boldsymbol{x}^T$ denote the transpose of vector $\boldsymbol{x}$. The Gram-Schmidt vectors are defined as follows:

$$\begin{aligned}
\boldsymbol{b}_1^* &= \boldsymbol{b}_1, \\
\boldsymbol{b}_i^* &= \boldsymbol{b}_i - \sum_{j=1}^{i-1} \mu_{ij}\boldsymbol{b}_j^*, \quad 2 \leq i \leq m, \quad \text{where} \\
\mu_{ij} &= \frac{\boldsymbol{b}_i^T\boldsymbol{b}_j^*}{\|\boldsymbol{b}_j^*\|^2}, \quad 1 \leq j < i \leq m.
\end{aligned}$$

For fixed $y \in (\frac{1}{4}, 1)$ we call $\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_m\}$ *y-reduced*, if

$$|\mu_{ij}| \leq \frac{1}{2}, \quad \text{for } 1 \leq j < i \leq m-1, \text{ and} \tag{5}$$

$$\|\boldsymbol{b}_i^* + \mu_{i,i-1}\boldsymbol{b}_{i-1}^*\|^2 \geq y\,\|\boldsymbol{b}_{i-1}^*\|^2, \quad \text{for } 1 < i \leq m-1\,. \tag{6}$$

Notice that, as $\boldsymbol{b}_1^*, \ldots, \boldsymbol{b}_m^*$ are pairwise orthogonal, Inequality (6) is satisfied if

$$\|\boldsymbol{b}_i^*\|^2 \geq y\|\boldsymbol{b}_{i-1}^*\|^2, \quad \text{for } 1 < i \leq m-1\,. \tag{7}$$

We will not describe the LLL-algorithm in detail, but just mention the two operations that are carried out by the algorithm. For $x \in \mathbb{R}^1$, let $\lfloor x \rceil$ denote the nearest integer to x. If Condition (5) is violated, i.e., $|\mu_{kj}| > 1/2$ for some $j < k$, then a *size reduction* is carried out by setting $\boldsymbol{b}_k := \boldsymbol{b}_k - \lfloor \mu_{kj} \rceil \boldsymbol{b}_j$. Notice that this operation will not change the Gram-Schmidt vector $\boldsymbol{b}_k^*$. If Condition (6) is violated for $i = j$, then vectors $\boldsymbol{b}_{j-1}$ and $\boldsymbol{b}_j$ are *interchanged*. This operation does affect several of the μ-values. Moreover, the new vector $\boldsymbol{b}_{j-1}^*$ will be the old vector $\boldsymbol{b}_j^* + \mu_{j,j-1}\boldsymbol{b}_{j-1}^*$.

For a given basis $\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_m\}$ of the lattice $L \subset \mathbb{R}^n$, define the matrix $\boldsymbol{B} = [\boldsymbol{b}_1 \cdots \boldsymbol{b}_m]$, such that the columns of $\boldsymbol{B}$ are given by the basis-vectors. Then $\boldsymbol{B}^T\boldsymbol{B}$ is an $m \times m$-matrix of full rank, and we can define the *determinant* of the lattice L as

$$d(L) = (\det(\boldsymbol{B}^T\boldsymbol{B}))^{1/2}\,. \tag{8}$$

It can be shown that this value is independent of the basis we choose for the lattice. Furthermore, we derive an expression of $d(L)$ in terms of the associated Gram-Schmidt orthogonalization.

Observation 1 *Given a basis $\boldsymbol{B} = [\boldsymbol{b}_1 \cdots \boldsymbol{b}_m]$ of a lattice $L \subset \mathbb{R}^n$ of rank m, and the associated Gram-Schmidt orthogonalization $\boldsymbol{B}^* = [\boldsymbol{b}_1^* \cdots \boldsymbol{b}_m^*]$, we have*

$$d(L) = \prod_{i=1}^{m} \|\boldsymbol{b}_i^*\|\,. \tag{9}$$

An explanation of how to derive Expression (9) can for instance be found in [10].

To every lattice L we can associate the *dual lattice*

$$L^\dagger = \{\boldsymbol{x} \in \operatorname{lin.span}(L) \mid \boldsymbol{x}^T\boldsymbol{y} \in \mathbb{Z} \text{ for all } \boldsymbol{y} \in L\}.$$

Notice that $L^{\dagger\dagger} = L$, and that

$$d(L^\dagger) = \frac{1}{d(L)}\,. \tag{10}$$

A subset $K \subseteq L$ is called a *pure sublattice* of L if $K = \operatorname{lin.span}(K) \cap L$. Let $K^\perp$ be the sublattice of $L^\dagger$ orthogonal to K, i.e., $K^\perp = \{\boldsymbol{x} \in L^\dagger \mid \boldsymbol{x}^T\boldsymbol{y} = 0 \text{ for all } \boldsymbol{y} \in K\}$.

Observation 2 *If K is a pure sublattice of L then $K^\perp$ is a pure sublattice of $L^\dagger$ and we have*

$$K^\perp = (L/K)^\dagger \tag{11}$$

and

$$d(L) = d(L/K) \cdot d(K)\,. \tag{12}$$

Suppose $L = \mathbb{Z}^n$. Then, by combining (12), (10), and (11) we obtain

$$d(K) = \frac{d(L)}{d(L/K)} = \frac{1}{d(L/K)} = d((L/K)^\dagger) = d(K^\perp)\,. \tag{13}$$

A more detailed account on this and much more can be found in, e.g., [14] and [15].

2.2 Some results for the kernel lattice of $\boldsymbol{a}$

In this subsection we consider a vector $\boldsymbol{a} \in \mathbb{Z}^n$ such that $\gcd(a_1, \ldots, a_n) = 1$.

The kernel lattice of $\boldsymbol{a}$ is the set $\ker_{\mathbb{Z}}(\boldsymbol{a}) := \{\boldsymbol{x} \in \mathbb{Z} \mid \boldsymbol{a}\boldsymbol{x} = 0\}$. The lattice $\ker_{\mathbb{Z}}(\boldsymbol{a})$ is a pure sublattice of $\mathbb{Z}^n$.

We first show in Lemma 1 that the lattice $\ker_{\mathbb{Z}}(\boldsymbol{a})$ has a basis of the following form:

$$\boldsymbol{Q} = \begin{pmatrix} x & x & \cdots & x \\ x & x & \cdots & x \\ 0 & x & \cdots & x \\ \vdots & 0 & \ddots & x \\ 0 & \cdots & 0 & x \end{pmatrix}, \tag{14}$$

where each ‘x’ denotes some integer number that may be different from zero.

Lemma 1. *The lattice* $\ker_{\mathbb{Z}}(\boldsymbol{a})$ *has a basis* $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{n-1}$ *of the following form:*

$$\mathbb{Z}\boldsymbol{b}_1 + \ldots + \mathbb{Z}\boldsymbol{b}_k = \ker_{\mathbb{Z}}(\boldsymbol{a}) \cap (\mathbb{Z}^{k+1} \times 0^{n-k-1}) \tag{15}$$

for any $1 \leq k \leq n-1$.

Proof. Write $c_i = \min\{|y_i| > 0 \mid \boldsymbol{y} \in \ker_{\mathbb{Z}}(\boldsymbol{a}), y_j = 0 \text{ for } j > i\}$, where $2 \leq i \leq n$. Note that the set we minimize over is not empty, because the vector $(-a_i, 0, \ldots, 0, a_1, 0, \ldots, 0)^T$, where a_1 appears in the ith position, is in $\ker_{\mathbb{Z}}(\boldsymbol{a})$ for any $i \in \{2, \ldots, n\}$. Now choose

$$\boldsymbol{b}_i \in \{\boldsymbol{x} \in \ker_{\mathbb{Z}}(\boldsymbol{a}) \mid x_{i+1} = c_{i+1}, x_j = 0 \text{ for } j > i+1\}. \tag{16}$$

To see that this is indeed a lattice-basis, let $\boldsymbol{z} \in \ker_{\mathbb{Z}}(\boldsymbol{a})$ and let k be the largest index of a non-zero coordinate of $\boldsymbol{z}$. Let $\boldsymbol{Q} = [\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{n-1}]$, where $\boldsymbol{b}_i$ satisfies (16).

We want to find $\boldsymbol{\lambda} \in \mathbb{Z}^{n-1}$ such that $\boldsymbol{z} = \boldsymbol{Q}\boldsymbol{\lambda}$. Observe that $\frac{z_k}{c_k}$ must be integer, because otherwise there is a $c' \in \mathbb{Z}$ such that $0 < |z_k - c'c_k| < c_k$, which contradicts the minimality of c_k. Therefore we may define $\lambda_{k-1} := \frac{z_k}{c_k}$.

Setting $\boldsymbol{z} = \boldsymbol{z} - \lambda_{k-1}\boldsymbol{b}_{k-1}$, this gives us a recursive construction for the integer coefficients $\lambda_1, \ldots, \lambda_{n-1}$ to express $\boldsymbol{z}$ in terms of our basis. □

One can additionally observe that if $\gcd(a_1, \ldots, a_i) = 1$ for some $1 \leq i \leq n$ then the last non-zero element of the basis vectors $\boldsymbol{b}_i, \ldots, \boldsymbol{b}_{n-1}$ is equal to ±1.

We will follow up on this idea in Section 4.

Let L_k be the sublattice given by the basis $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k$ as described in Lemma 1, for $1 \leq k \leq m$. Then we have $L_1 \subseteq L_2 \subseteq \cdots \subseteq L_{n-1} = \ker_{\mathbb{Z}}(\boldsymbol{a})$ and $d(L_k) = \prod_{i=1}^{k} \|\boldsymbol{b}_i^*\|$. Also, because of the specific structure of the basis, we can express L_k as

$$L_k = \{\boldsymbol{x} \in \mathbb{Z}^n \mid (a_1, \ldots, a_{k+1}, 0, \ldots, 0)\boldsymbol{x} = 0, x_j = 0, k+2 \leq j \leq n\}.$$

We can extend the above observations to conclude the following:

Lemma 2. *Let $L_1, \ldots, L_{n-1}$ be given as above and let $k \in \{1, \ldots, n-1\}$. If* $\gcd(a_1, \ldots, a_{k+1}) = 1$*, then*

$$d(L_k) = \sqrt{\sum_{i=1}^{k+1} a_i^2}, \tag{17}$$

and thus we get in particular

$$\|\boldsymbol{b}_k^*\|^2 = \frac{\sum_{i=1}^{k+1} a_i^2}{\sum_{i=1}^{k} a_i^2}. \tag{18}$$

Proof. Observe that $(a_1, \ldots, a_{k+1}, 0, \ldots, 0)^T$ and the unit vectors $\boldsymbol{e}_j$, with $k+2 \leq j \leq n$, are an orthogonal basis of $L_k^\perp$. Using (9) and the fact that $d(K) = d(K^\perp)$ for pure sublattices of $\mathbb{Z}^n$ (see (13)), we get (17).

Equation (18) follows from (9) in combination with (17) for L_k and L_{k-1}. □

3 Probabilistic analysis

Here we present the main result of the paper, namely a bound on the probability that the LLL-algorithm will perform a basis vector interchange after basis vector $\boldsymbol{b}_k$ is considered. We assume that the elements a_i of the vector $\boldsymbol{a}$ are drawn independently and uniformly at random from an interval $[l, \ldots, u] := [l, u] \cap \mathbb{Z}$, where $0 < l < u$, and that the starting basis of $\ker_{\mathbb{Z}}(\boldsymbol{a})$ is a basis of the structure given in Lemma 1. Recall from Subsection 2.1 that if, for given reduction factor $y \in (\frac{1}{4}, 1)$,

$$\|\boldsymbol{b}_{i+1}^*\|^2 \geq y\|\boldsymbol{b}_i^*\|^2, \quad \text{for } 1 \leq i < n-1\,,$$

then the LLL-algorithm will not interchange basis vectors $\boldsymbol{b}_i$ and $\boldsymbol{b}_{i+1}$.

We will prove the following result:

Theorem 1. *Let $y \in (\frac{1}{4}, 1)$ be fixed. Then, for k large enough, we get*

$$\Pr\left(\frac{\|\boldsymbol{b}_{k+1}^*\|^2}{\|\boldsymbol{b}_k^*\|^2} \leq y\right) \quad \leq \quad e^{-c(k+1)^2} + 2^{-(k+1)/2}\,, \tag{19}$$

where $c > 0$ depends on $u, l,$ and y.

We provide explicit bounds on c and when k is large enough. To increase accessibility to the proof, we build our result from several lemmas. We start by noticing that for any $1 \leq k < n-1$

$$\begin{aligned}\Pr\left(\|\boldsymbol{b}_{k+1}^*\|^2 < y\|\boldsymbol{b}_k^*\|^2\right) \leq\ & \Pr\left(\|\boldsymbol{b}_{k+1}^*\|^2 < y\|\boldsymbol{b}_k^*\|^2 \mid \gcd(a_1, \ldots, a_{k+1}) = 1\right) \\ & + \Pr(\gcd(a_1, \ldots, a_{k+1}) > 1),\end{aligned} \tag{20}$$

and hence we can bound the two terms separately. The last one can be bounded in the following way:

Lemma 3. *Let $a_1, \dots, a_n$ be chosen independently and uniformly at random from $[l, \dots, u]$ for some integers $0 < l < u$, and let l and u be fixed. Then*

$$\Pr(\gcd(a_1, \dots, a_{k+1}) > 1) \leq \left(\frac{1}{2}\right)^{(k+1)/2}$$

for any $k \geq \frac{\log_2\left(\left\lfloor \frac{u}{2} \right\rfloor + 1\right)}{\log_2\left(\frac{u-l+1}{u-l+2}\right) + \frac{1}{2}}$.

Next, for given reduction factor y, we want to derive a bound on the first term of Expression (20), i.e.:

$$\Pr\left(\frac{\|\boldsymbol{b}_{k+1}^*\|^2}{\|\boldsymbol{b}_k^*\|^2} < y \mid \gcd(a_1, \dots, a_{k+1}) = 1\right).$$

Showing that the ratio between $\|\boldsymbol{b}_{k+1}^*\|^2$ and $\|\boldsymbol{b}_k^*\|^2$ behaves the way we suspect is not straightforward as the two quantities are not independent. To estimate the mean of this ratio we use a result by Pittenger [19], which we state below in a form that is adapted to our situation.

Theorem 2 ([19], adapted). *Let X be a random variable on some positive domain. Choose $c > 0$ such that $X - c \geq 0$ and define $\mu = \mathbb{E}[X]$ and $\sigma^2 = Var(X)$. Then*

$$\begin{aligned} \frac{1}{\mu} &\leq \mathbb{E}\left[\frac{1}{X}\right] \\ &\leq \frac{\mu^3 c - 3\mu^2 c^2 + 3\mu c^3 - c^4 + \sigma^2\mu^2 - \sigma^2\mu c + \sigma^4}{\mu^4 c - 3\mu^3 c^2 + 3\mu^2 c^3 - \mu c^4 + 2\sigma^2\mu^2 c - 3\sigma^2\mu c^2 + \sigma^2 c^3 + \sigma^4 c}. \end{aligned} \tag{21}$$

For convenience of notation we define $X_k := \sum_{i=1}^k a_i^2$. We first estimate the following mean.

Lemma 4. *Let $a_1, \dots, a_n$ be chosen independently and uniformly at random from $[l, \dots, u]$ for some integers $0 < l < u$, let $\boldsymbol{b}_1, \dots, \boldsymbol{b}_{n-1}$ be given as in Lemma 1, and let $1 < k < n$.*

If $\gcd(a_1, \dots, a_{k+1}) = 1$, there exists a function $f(k) \in \Theta(\frac{1}{k^2})$ such that

$$1 + \frac{1}{k} \leq \mathbb{E}\left[\|\boldsymbol{b}_k^*\|^2\right] \leq 1 + \frac{1}{k} + f(k), \tag{22}$$

and we can give an explicit expression for $f(k)$.

Note that using Theorem 2, we can compute an explicit upper bound in (22). We present this upper bound in the complete version of our paper.

Lemma 5. *Let $a_1, \dots, a_n$ be chosen independently and uniformly at random from $[l, \dots, u]$ for some integers $0 < l < u$. Then for any $1 \leq k < n - 1$ with $\gcd(a_1, \dots, a_{k+1}) = 1$ we get*

$$\left|1 - \mathbb{E}\left[\|\boldsymbol{b}_{k+1}^*\|^2 / \|\boldsymbol{b}_k^*\|^2\right]\right| = O\left(\frac{1}{k}\right). \tag{23}$$

As with Lemma 4, we give explicit upper and lower bounds in the complete version of our paper.

Returning to Inequality (20), we will in fact only need the lower bound for $\mathbb{E}\left[\|\boldsymbol{b}_{k+1}^*\|^2/\|\boldsymbol{b}_k^*\|^2\right]$, to see that for any given reduction factor y we can find a $k(y)$ such that the mean is larger than y for any $k \geq k(y)$. More precisely:

Corollary 1. *Let $a_1, \ldots, a_n$ be chosen independently and uniformly at random from $[l, \ldots, u]$ for some integers $0 < l < u$, and let $y \in (1/4, 1)$ be fixed. Define $\hat{\mu} := \mathbb{E}[a_i^2]$ and $\hat{\sigma}^2 := Var(a_i^2)$.*

Suppose $k \leq n$ is given, and $\gcd(a_1, \ldots, a_{k+1}) = 1$. If k satisfies

$$1 - \frac{u^2 - \hat{\mu}}{(k+1)\hat{\mu}} - \frac{u^2\hat{\mu}}{(k+1)^2\hat{\mu}^2 + (k+1)\hat{\sigma}^2} > y, \tag{24}$$

then $\mathbb{E}\left[\frac{\|\boldsymbol{b}_{k+1}^*\|^2}{\|\boldsymbol{b}_k^*\|^2}\right] > y$.

Note that (24) can be solved explicitly for $k+1$, giving us a lower bound on k. We omit this calculation here as the solution is long and does not seem illuminating as to what size is sufficient for k. We will give some examples for given l, u, and y in Section 4.

If we can now also control the probability of $\|\boldsymbol{b}_{k+1}^*\|/\|\boldsymbol{b}_k^*\|$ deviating by more than a small amount from its mean for given $\boldsymbol{a}$, we have found a bound on the first term on the right in (20). For this we apply the inequality of Azuma-Hoeffding (cf. [4, 8]):

Let $Z_1, \ldots, Z_N$ be independent random variables, where Z_i takes values in the space Λ_i, and let $f : \prod_{i=1}^N \Lambda_i \to \mathbb{R}$. Define the following Lipschitz condition for the numbers $c_1, \ldots, c_N$:

(L) If the vectors $\boldsymbol{z}, \boldsymbol{z}' \in \prod_{i=1}^N \Lambda_i$ differ only in the jth coordinate, then $|f(\boldsymbol{z}) - f(\boldsymbol{z}')| \leq c_j$, for $j = 1, \ldots, N$.

Theorem 3 (see [9]). *If f is measurable and satisfies (L), then the random variable $X = f(Z_1, \ldots, Z_N)$ satisfies, for any $t \geq 0$,*

$$\begin{aligned} \Pr\left(X \geq \mathbb{E}[X] + t\right) \leq & \quad e^{\frac{-2t^2}{\sum_{i=1}^N c_i^2}} \text{ and} \\ \Pr\left(X \leq \mathbb{E}[X] - t\right) \leq & \quad e^{\frac{-2t^2}{\sum_{i=1}^N c_i^2}}. \end{aligned} \tag{25}$$

Thus, we indeed have a bound on the probability that a random variable satisfying (L) will deviate more than a little bit from its mean. Note that the bound gets stronger if we find small c_i and choose t large.

As with Lemma 5, we will ultimately just need one of the bounds, in this case (25).

Applied to our situation, we obtain the following result.

Corollary 2. *Let $a_1, \ldots, a_n$ be chosen independently and uniformly at random from $[l, \ldots, u]$ for some integers $0 < l < u$, and let $y \in (1/4, 1)$ be fixed.*

Suppose $k < n$ is given, and $\gcd(a_1, \ldots, a_{k+1}) = 1$. If k satisfies (24), then

$$\Pr\left(\frac{\|\boldsymbol{b}_{k+1}^*\|^2}{\|\boldsymbol{b}_k^*\|^2} \leq y\right) \quad \leq \quad e^{-t^2(k+1)^2\hat{c}}, \tag{26}$$

where $\hat{c} > 0$ depends on u and l, and $t > 0$ depends on u, l, and y.

To summarize, we proved in Lemma 3 and in Corollary 2 that for fixed reduction factor $y \in (1/4,\ 1)$, and for fixed l, u the following holds:

$$\Pr(\gcd(a_1, \ldots, a_{k+1}) > 1) \leq \left(\frac{1}{2}\right)^{(k+1)/2} \quad \text{for any } k \geq \frac{\log_2\left(\left\lfloor \frac{u}{2} \right\rfloor + 1\right)}{\log_2\left(\frac{u-l+1}{u-l+2}\right) + \frac{1}{2}} \tag{27}$$

and,

$$\Pr\left(\|\boldsymbol{b}_{k+1}^*\|^2 < y\|\boldsymbol{b}_k^*\|^2 \mid \gcd(a_1, \ldots, a_{k+1}) = 1\right) \quad \leq \quad e^{-t^2(k+1)^2\hat{c}}, \tag{28}$$

where $\hat{c} > 0$ depends on u and l, and $t > 0$ depends on u, l, and y. Adding the right-hand sides of Inequalities (27) and (28) yields the upper bound on $\Pr\left(\|\boldsymbol{b}_{k+1}^*\|^2 / \|\boldsymbol{b}_k^*\|^2 \leq y\right)$ as stated in Theorem 1.

4 Discussion and computations

If we again look at a basis $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k$ that is obtained by applying the LLL reduction algorithm to an input basis of the format described in Lemma 1 in Subsection 2.2, we showed that for not too small k it will most likely have the following structure:

$$\left(\begin{array}{c|c} X_1 & X_2 \\ \hline \mathbf{0} & X_3 \end{array}\right).$$

The dimension of the submatrices X_1, X_2 and X_3 are $(k+1) \times k$, $(k+1) \times (n-(k+1))$, and $(n-(k+1)) \times (n-(k+1))$ respectively. All the elements of X_1 and X_2 may be non-zero, and X_3 is upper triangular.

In our computations, however, we see even more structure in the reduced basis, as discussed in the introduction. More precisely, we observe a reduced basis of the following form:

$$\left(\begin{array}{c|c} X_1 & \bar{X}_2 \\ \hline \mathbf{0} & I \end{array}\right), \tag{29}$$

that is, $X_3 = I$. So, a remaining question to address is why this is the case. We pointed out in Subsection 2.2 that if $\gcd(a_1, \ldots, a_{k+1}) = 1$, then it follows from the proof of Lemma 1 that the last nonzero element in each of the columns $\boldsymbol{b}_{k+1}, \ldots, \boldsymbol{b}_{n-1}$ must be ± 1. Therefore we know that the first column of X_3 is $(1, 0, \ldots, 0)^T$. The second column of X_3 is $(x, 1, 0, \ldots, 0)^T$, and so on. Here, again, x just denotes that the element may be non-zero. So, by subtracting x times vector $\boldsymbol{b}_{k+1}$ from vector $\boldsymbol{b}_{k+2}$ yields a unit column $(0, 1, 0, \ldots, 0)^T$ as the second column of X_3. This procedure can now be repeated for the remaining basis vectors to produce $X_3 = I$. Notice that these operations are elementary column operations.

Observation 3 *If we apply the above column operations to the basis given in Lemma 1, then every part of the analysis where we assumed the basis to be given as in Lemma 1 also works for this new lattice basis.*

So, indeed, $\ker_{\mathbb{Z}}(\boldsymbol{a})$ has a basis of the structure given in (29), and we observe in our computational experiments that such a basis is y-reduced if the input vector $\boldsymbol{a}$ satisfies the assumptions given in the beginning of Section 3. Here we give qualitative arguments for why this is the case.

Suppose that the elementary column operations performed to obtain $X_3 = I$ yields a basis that is not size reduced. Then we can add any linear integer combination of the first k basis vectors to any of the last $n - (k + 1)$ vectors without destroying the identity matrix structure of submatrix X_3, since the first k vectors have zeros as the last $n - (k + 1)$ elements. These elementary column operations can be viewed as size reductions. If we consider the first k basis vectors we empirically observe that the absolute values of the non-zero elements (i.e., elements in submatrix X_1) are small, and that the vectors are almost orthogonal since they are reduced. Since all a_i-elements are positive, each basis vector has a mixture of positive, negative and zero elements. Apparently, once these size reductions are done, the basis is reduced, i.e., no further swaps are needed. This is in line with the results presented in Subsection 3 that the expected length of the Gram-Schmidt vectors $\boldsymbol{b}_k^*$ becomes arbitrarily close to one with increasing values of k, see also reduction Condition (7).

In Table 1 we give an upper bound on $\Pr(\gcd(a_1, \ldots, a_{k+1}) > 1)$ for k greater than or equal to the value given in the table. This probability is computed according to Lemma 3 for the intervals $[l, \ldots, u] = [100, \ldots, 1,000]$ and $[l, \ldots, u] = [15,000, \ldots, 150,000]$. That is, for the interval $[l, \ldots, u] = [100, \ldots, 1,000]$, the probability that $\gcd(a_1, \ldots, a_{k+1}) > 1$ is less than or equal to 0.0014 for $k \geq 19$. Notice that this value of k is only depending on l and u, and not on n. In the table we also give the value of $k(y)$ for reduction factor $y = 95/100$ such that $\mathbb{E}\left[\|\boldsymbol{b}_{k+1}^*\|^2/\|\boldsymbol{b}_k^*\|^2\right] > y$ for all $k \geq k(y)$. The values given in the table are very close to the values we observe empirically.

A comprehensive computational study for single- and multi-row instances is presented in the complete version of our paper.

Table 1. Column two gives an upper bound on $\Pr(\gcd(a_1, \ldots, a_{k+1}) > 1)$ for k greater than or equal to the value given in column 3, cf. Lemma 3. In the fourth column we give the value of $k(y)$ for reduction factor $y = 95/100$, such that $\mathbb{E}\left[\|\boldsymbol{b}_{k+1}^*\|^2/\|\boldsymbol{b}_k^*\|^2\right] > y$ for all $k \geq k(y)$.

Interval	Probability $\leq$	$k \geq$	$k(y)$
$[100, \ldots, 1,000]$	0.0014	19	36
$[15,000, \ldots, 150,000]$	0.000008	34	36

To summarize, we have observed empirically that for larger instances, only relatively few of the $\boldsymbol{x}$-variables have a non-trivial translation into $\boldsymbol{\lambda}$-variables. This is well in line with the theoretical result reported in Table 1 that the expected value of $\|\boldsymbol{b}_{k+1}^*\|^2/\|\boldsymbol{b}_k^*\|^2$ is greater than the reduction factor for all $k \geq 36$ for both of the considered intervals. Yet, we observe that if we solve the instances using Reformulation (3) rather than the original formulation (1), the number of branch-and-bound nodes needed in $\boldsymbol{\lambda}$-space could be one to two orders of magnitude smaller than in the original space. Thus, there is a computationally important structure in the $\boldsymbol{\lambda}$-space. But this structure is not arbitrarily "spread", but contained in a limited subset of the variables.

Suppose now that a row $\boldsymbol{ax} = b$ is part of a larger problem formulation, and that we expect this row to be important in the formulation in the sense of obtaining a good branching direction or a useful cut. If we wish to obtain this information through the lattice reformulation, then we need to be careful in indexing the $\boldsymbol{x}$-variables appropriately.

Acknowledgement

We wish to acknowledge the fruitful discussion with Andrea Lodi and Laurence Wolsey that lead to the question addressed in this contribution. We also want to thank Hendrik Lenstra for his helpful suggestions.

References

1. Aardal, K., Hurkens, C.A.J., Lenstra, A.K.: Solving a system of linear Diophantine equations with lower and upper bounds on the variables. Mathematics of Operations Research 25(3), 427–442 (2000)
2. Aardal, K., Lenstra, A.K.: Hard equality constrained integer knapsacks. Mathematics of Operations Research 29(3), 724–738 (2004). Erratum: Mathematics of Operations Research 31(4), 846 (2006)
3. Aardal, K., Wolsey, L.A.: Lattice based extended formulations for integer linear equality systems. Mathematical Programming 121, 337–352 (2010)
4. Azuma, K.: Weighted sums of certain dependent random variables. Tôhoku Mathematical Journal 19(3), 357–367 (1967)
5. Cook, W., Rutherford, T., Scarf, H.E., Shallcross, D.: An implementation of the generalized basis reduction algorithm for integer programming. ORSA Journal on Computing 5, 206–212 (1993)
6. Cornuéjols, G., Dawande, M.: A class of hard small 0-1 programs. INFORMS Journal on Computing 11, 205–210 (1999)
7. Cornuéjols, G., Urbaniak, R., Weismantel, R., Wolsey, L.A.: Decomposition of integer programs and of generating sets. In: Burkard, R.E., Woeginger, G.J. (eds.) Algorithms – ESA '97. LNCS, vol. 1284, pp. 92–103. Springer, Heidelberg (1997)
8. Hoeffding, W.: Probability inequalities for sums of bounded random variables. Journal of the American Statistical Association 58, 13–30 (1963)
9. Janson, S.: On concentration of probability. In: Contemporary combinatorics, volume 10 of *Bolyai Soc. Math. Stud.*, pp. 289–301. János Bolyai Math. Soc., Budapest, (2002)

10. Kannan, R.: Algorithmic geometry of numbers. In: Annual review of computer science 2, pp. 231–267. Annual Reviews, Palo Alto, CA (1987)
11. Krishnamoorthy B., Pataki, G.: Column basis reduction and decomposable knapsack problems. Discrete Optimization 6(3), 242–270 (2009)
12. Lenstra, A.K., Lenstra, Jr., H.W., Lovász, L.: Factoring polynomials with rational coefficients. Mathematische Annalen 261(4), 515–534 (1982)
13. Lenstra, Jr., H.W.: Integer programming with a fixed number of variables. Mathematics of Operations Research 8(4), 538–548 (1983)
14. Lenstra, Jr., H.W.: Flags and lattice basis reduction. In: European Congress of Mathematics, Vol. I (Barcelona, 2000), volume 201 of Progress in Mathematics, pp. 37–51. Birkhäuser, Basel (2001)
15. Lenstra, Jr., H.W.: Lattices. In: Algorithmic number theory: lattices, number fields, curves and cryptography, volume 44 of Mathematical Science Research Institute Publications, pp. 127–181. Cambridge University Press, Cambridge (2008)
16. Louveaux Q., Wolsey, L.A.: Combining problem structure with basis reduction to solve a class of hard integer programs. Mathematics of Operations Research 27(3), 470–484 (2002)
17. Lovász, L., Scarf, H.E.: The generalized basis reduction algorithm. Mathematics of Operations Research 17, 751–764 (1992)
18. Mehrotra, S., Li, Z.: Branching on hyperplane methods for mixed integer linear and convex programming using adjoint lattices. Journal of Global Optimization 49(4), 623–649 (2011)
19. Pittenger, A.O.: Sharp mean-variance bounds for Jensen-type inequalities. Statistics & Probabability Letters 10(2), 91–94 (1990)