# Proximity bounds for random integer programs

Marcel Celaya* and Martin Henk

Technische Universität Berlin
Institut für Mathematik, Sekr. MA4-1
Straße des 17 Juni 136
D-10623 Berlin

{henk,celaya}@math.tu-berlin.de

No Institute Given

**Abstract.** We study proximity bounds within a natural model of random integer programs of the type $\max \boldsymbol{c}^\top \boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \in \mathbb{Z}_{\geq 0}$, where $\boldsymbol{A} \in \mathbb{Z}^{m\times n}$ of rank $m$, $\boldsymbol{b} \in \mathbb{Z}^m$ and $\boldsymbol{c} \in \mathbb{Z}^n$. We prove that (up to a constant depending on $n$) the proximity is "generally" bounded by $\Delta_m(\boldsymbol{A})^{1/(n-m)}$, where $\Delta_m(\boldsymbol{A})$ is the maximal absolute value of an $m\times m$ subdeterminant of $\boldsymbol{A}$. This is significantly better than the best deterministic bounds which are linear in $\Delta_m(\boldsymbol{A})$.

## 1 Introduction

Given an linear program of the form

$$\max\ \boldsymbol{c}^\top \boldsymbol{x}\ :\ \begin{aligned}\boldsymbol{A}\boldsymbol{x} &= \boldsymbol{b}\\ \boldsymbol{x} &\geq \boldsymbol{0},\end{aligned} \tag{1}$$

where $\boldsymbol{A}$ is a full-row-rank $m \times n$ integral matrix, $\boldsymbol{b} \in \mathbb{Z}^m$, and $\boldsymbol{c} \in \mathbb{Z}^n$, the proximity problem seeks to understand how far away an optimal feasible solution $\boldsymbol{x}^*$ can be to a nearby feasible integer solution $\boldsymbol{z}^*$. Assuming the feasible region has at least one such integral point, bounds for proximity are typically given in terms of the largest possible absolute value $\Delta_m(\boldsymbol{A})$ of any $m\times m$ subdeterminant of $\boldsymbol{A}$. This is a well-studied problem which goes back to the classic Cook et al. result [4] bounding the proximity of the dual of (1). See, for instance, the recent works of Eisenbrand and Weismantel [5] and of Aliev, Henk, and Oertel [1] and the references therein.

In this manuscript, we would like to understand the worst-possible proximity, which we denote by $\operatorname{dist}(\boldsymbol{A})$, over all choices of $\boldsymbol{b}$ and $\boldsymbol{c}$, when the matrix $\boldsymbol{A}$ is chosen *randomly*. The model of randomness we consider is the following: we choose the matrix $\boldsymbol{A}$ up to left-multiplication by unimodular matrices, and we choose

* The first author was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy — The Berlin Mathematics Research Center MATH+(EXC-2046/1, project ID: 390685689).

$\boldsymbol{A}$ uniformly at random subject to the condition that $\Delta(\boldsymbol{A}) := \sqrt{\det \boldsymbol{A}\boldsymbol{A}^\top}$ is at most some large, fixed integer $T$. This is a natural model to study from a geometric point of view, as $\Delta(\boldsymbol{A})$ is the determinant of the lattice of integer points in the kernel of $\boldsymbol{A}$. This is also the model considered by Aliev and Henk in [2], in their investigation of diagonal Frobenius numbers.

Our main result concerns not $\operatorname{dist}(\boldsymbol{A})$ but rather a related random variable we denote by $\operatorname{dist}^*(\boldsymbol{A})$. This is an asymptotic version of $\operatorname{dist}(\boldsymbol{A})$ that further imposes some mild restrictions on $\boldsymbol{b}$. Our main result is that it satisfies the following Markov-type inequality:

$$\mathbf{P}\left(\operatorname{dist}^*(\boldsymbol{A}) > t\Delta(\boldsymbol{A})^{1/(n-m)}\right) \ll t^{-2/3}. \tag{2}$$

Here $\ll$ means less than, up to constants which only depend on the dimension. In particular, this shows that proximity generally depends only on $\Delta^{1/(n-m)}$ in the random setting, for "almost all" choices of $\boldsymbol{b}$ in a certain precise sense. This is significantly better than the linear dependency on $\Delta_m$ in the deterministic case, that is known to be tight [1, Theorem 1]. A similar result, with a slightly different random model, was obtained in [1] the so-called knapsack scenario, where $m = 1$. We also mention recent work of Oertel, Paat, and Weismantel in [8], which considers a random model that allows $\boldsymbol{b}$ to vary but keeps $\boldsymbol{A}$ fixed.

The proof of this result combines ideas of [2] and [1] using facts from the geometry of numbers, some results of Schmidt from [9] on random sublattices of $\mathbb{Z}^n$ of fixed dimension, and computations of the measure of certain distinguished regions of the real Grassmannian $\operatorname{Gr}(n, d)$ of $d$-dimensional subspaces of $\mathbb{R}^n$. The idea is two-fold. First, we use the results of Schmidt to relate the discrete measure in our model to the continuous $O(n)$-invariant probability measure $\nu$ of $\operatorname{Gr}(n, d)$. We show that there are essentially two distinct "bad" regions of $\operatorname{Gr}(n, d)$, both parameterized by $t$, in which $\operatorname{dist}^*(\boldsymbol{A})$ could be large, but whose measure with respect to $\nu$ gets smaller as $t$ gets larger.

We remark that the exponent of $-2/3$ is mainly an artifact of the proof, and we expect that it can be further improved. The problem of finding an inequality analogous to (2) for $\operatorname{dist}(\boldsymbol{A})$ is more challenging and remains open, as the polyhedral combinatorics of (1) may interfere with our analysis.

## 2 Main result and notation

### 2.1 Notation

Throughout this manuscript we assume fixed positive integers $d, m, n$ such that $n = m+d$. For a subset $\sigma \subseteq [n]$ and $\boldsymbol{x} \in \mathbb{R}^n$, we let $\boldsymbol{x}_\sigma$ denote the vector obtained by orthogonally projecting $\boldsymbol{x}$ onto the coordinates indexed by $\sigma$. Similarly, if $\boldsymbol{A}$ is a matrix, then we denote by $\boldsymbol{A}_\sigma$ the submatrix of $\boldsymbol{A}$ whose columns are those indexed by $\sigma$. In particular, if $k \in [n]$ then $\boldsymbol{A}_k$ denotes the corresponding column of $\boldsymbol{A}$. If $\boldsymbol{A}_\sigma$ is an invertible square matrix we say $\sigma$ is a *basis* of $\boldsymbol{A}$. We denote the complement of $\sigma$ by $\bar{\sigma} := [n]\backslash\sigma$. Given a $d$-dimensional subspace $L \subseteq \mathbb{R}^n$, the $m$-dimensional orthogonal complement of $L$ is denoted by $L^\perp$. If $\Lambda \subset \mathbb{R}^n$, let $\Lambda_\mathbb{R}$ denote the linear subspace of $\mathbb{R}^n$ spanned by $\Lambda$.

### 2.2 Definition of dist($\boldsymbol{A}$)

Given a full-rank matrix $\boldsymbol{A} \in \mathbb{Z}^{m\times n}$ and vector $\boldsymbol{b} \in \mathbb{Z}^m$, we define the polyhedron

$$\mathcal{P}(\boldsymbol{A}, \boldsymbol{b}) := \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},\ \boldsymbol{x} \geq \boldsymbol{0}\}\,.$$

Given a vertex $\boldsymbol{x}^*$ of this polyhedron, we define

$$\operatorname{dist}(\boldsymbol{A}, \boldsymbol{b}, \boldsymbol{x}^*) := \inf_{\boldsymbol{z}^* \in \mathbb{Z}^n \cap \mathcal{P}(\boldsymbol{A},\boldsymbol{b})} \|\boldsymbol{x}^* - \boldsymbol{z}^*\|_2\,.$$

We then define the worst-case distance over all right-hand-side vectors $\boldsymbol{b} \in \mathbb{Z}^m$ for which $\mathcal{P}(\boldsymbol{A}, \boldsymbol{b})$ is nonempty, and over all vertices $\boldsymbol{x}^*$ of $\mathcal{P}(\boldsymbol{A}, \boldsymbol{b})$:

$$\operatorname{dist}(\boldsymbol{A}) := \sup_{\boldsymbol{b}} \sup_{\boldsymbol{x}^*} \operatorname{dist}(\boldsymbol{A}, \boldsymbol{b}, \boldsymbol{x}^*)\,. \tag{3}$$

This definition has the disadvantage that it is stated in terms of the matrix $\boldsymbol{A}$. Since we may replace $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ with $\boldsymbol{U}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{U}\boldsymbol{b}$ for any unimodular $m\times m$ matrix $\boldsymbol{U}$, it is not so clear from this formulation how to define our random model. This motivates an alternative, more geometric definition of dist($\boldsymbol{A}$) which we now state.

### 2.3 Definition of dist($\Lambda$)

Suppose instead we start with a *primitive* $d$-dimensional sublattice $\Lambda$ of $\mathbb{Z}^n$, meaning that $\Lambda = \Lambda_{\mathbb{R}} \cap \mathbb{Z}^n$. Let $\boldsymbol{A} \in \mathbb{Z}^{m\times n}$ be any integral matrix for which the rows of $\boldsymbol{A}$ form a basis of $\Lambda_{\mathbb{R}}^{\perp} \cap \mathbb{Z}^n$. Suppose $\sigma$ is basis of $\boldsymbol{A}$, and $\boldsymbol{x}^*$ is a vector lying in the semigroup

$$\mathcal{S}_\sigma := \left\{\boldsymbol{x} \in \mathbb{Q}^n : \boldsymbol{x}_\sigma \in \boldsymbol{A}_\sigma^{-1}\mathbb{Z}^m,\ \boldsymbol{x}_{\bar{\sigma}} = \boldsymbol{0},\ \boldsymbol{x}_\sigma \geq \boldsymbol{0}\right\}.$$

Then we may define

$$\operatorname{dist}(\Lambda, \sigma, \boldsymbol{x}^*) := \operatorname{dist}(\boldsymbol{A}, \boldsymbol{b}, \boldsymbol{x}^*)\,, \tag{4}$$

where $\boldsymbol{b} := \boldsymbol{A}\boldsymbol{x}^*$, and

$$\operatorname{dist}(\Lambda) := \sup_{\sigma} \sup_{\boldsymbol{x}^*} \operatorname{dist}(\Lambda, \sigma, \boldsymbol{x}^*)\,, \tag{5}$$

where the supremum is taken over all bases $\sigma$ of $\boldsymbol{A}$ and elements $\boldsymbol{x}^* \in \mathcal{S}_\sigma$. Importantly, note that definitions (4) and (5) do not depent on the choice of $\boldsymbol{A}$, but only on the data $(\Lambda, \sigma, \boldsymbol{x}^*)$ and $\Lambda$, respectively.

One easily checks that the definitions of (3) and (5) agree. Nevertheless, the advantage of definition (5) is that there are only finitely many $d$-dimensional sublattices $\Lambda$ of $\mathbb{Z}^n$ whose determinant $\Delta(\Lambda) := \Delta(\boldsymbol{A})$ is at most some fixed positive integer $T$. Thus, we may consider the uniform distribution over these bounded-determinant lattices.

### 2.4 An asymptotic version of dist($\Lambda$)

Again assume the rows of $\boldsymbol{A}$ form a basis of $\Lambda_{\mathbb{R}}^{\perp} \cap \mathbb{Z}^n$. We next consider a related version of $\operatorname{dist}(\boldsymbol{A})$, or equivalently $\operatorname{dist}(\Lambda)$. Let $B_2^n \subset \mathbb{R}^n$ denote the $n$-dimensional Euclidean ball. Define the vector $\boldsymbol{w} \in \mathbb{R}^n$ as follows: for each $i \in [n]$, set

$$\boldsymbol{w}_i := \sqrt{1 - \boldsymbol{A}_i^{\top} (\boldsymbol{A}\boldsymbol{A}^{\top})^{-1} \boldsymbol{A}_i}. \tag{6}$$

This vector $\boldsymbol{w}$ measures, for each $i \in [n]$, the largest possible value of $\boldsymbol{x}_i$ for any $\boldsymbol{x} \in B_2^n \cap \Lambda_{\mathbb{R}}$. Denote by $\mu := \mu(\Lambda, B_2^n)$ the covering radius of $B_2^n$ with respect to $\Lambda$. That is,

$$\mu := \inf \{t > 0 : \Lambda + tB_2^n \text{ contains } \Lambda_{\mathbb{R}}\}.$$

If $\sigma$ is a basis of $\boldsymbol{A}$ then define the following subsemigroup of $\mathcal{S}_\sigma$:

$$\mathcal{S}_\sigma^* := \left\{\boldsymbol{x} \in \mathcal{S}_\sigma : \boldsymbol{x}_\sigma \geq \mu \boldsymbol{w}_\sigma + \boldsymbol{A}_\sigma^{-1} \boldsymbol{A}_{\bar{\sigma}} \boldsymbol{w}_{\bar{\sigma}}\right\}.$$

The next proposition shows that if we further restrict $\boldsymbol{x}^*$ so that it can only lie in $\mathcal{S}_\sigma^*$, then we can guarantee that $\mathcal{P}(\boldsymbol{A}, \boldsymbol{b})$ contains an integral point reasonably close to $\boldsymbol{x}^*$. We prove it in Section 5.

**Proposition 1.** *For a basis $\sigma$ of $\boldsymbol{A}$ and $\boldsymbol{x}^* \in \mathcal{S}_\sigma^*$, let $\boldsymbol{b} = \boldsymbol{A}\boldsymbol{x}^*$. Then $\mathcal{P}(\boldsymbol{A}, \boldsymbol{b})$ contains a translate of the scaled ball $\mu \cdot (B_2^n \cap \Lambda_{\mathbb{R}})$, which in turn contains an integral vector.*

Now set

$$\operatorname{dist}^*(\Lambda) := \sup_{\sigma} \sup_{\boldsymbol{x}^*} \operatorname{dist}(\Lambda, \sigma, \boldsymbol{x}^*), \tag{7}$$

where the supremum is taken over all bases $\sigma$ of $\boldsymbol{A}$ and elements $\boldsymbol{x}^*$ of the semigroup $\mathcal{S}_\sigma^*$.

### 2.5 Main result

We are now ready to state the main theorem.

**Theorem 1.** *For $T \gg 1$, let $\Lambda$ be a primitive sublattice of $\mathbb{Z}^n$ of dimension $d$ and determinant at most $T$, chosen uniformly at random. Then for all $t > 1$,*

$$\mathbf{P}\left(\operatorname{dist}^*(\Lambda) > t\left(\Delta(\Lambda)\right)^{1/d}\right) \ll t^{-2/3}.$$

## 3 A theorem of Schmidt

In this section we state a result that is fundamental to the proof, which follows from the results of Schmidt in [9]. We continue with our assumption that $d = n - m$. Let $\operatorname{Gr}(d, n)$ denote the set of $d$-dimensional subspaces of $\mathbb{R}^n$. Let $\nu$ denote the unique $O(n)$-invariant probability measure on the real Grassmannian $\operatorname{Gr}(d, n)$.

**Definition 1 ([9, p. 40]).** *A set $\xi \subset \mathrm{Gr}(d,n)$ is* Jordan measurable *if for all $\varepsilon > 0$ there exists continuous functions $f_1 \leq \mathbf{1}_\xi \leq f_2$ such that*

$$\int (f_2 - f_1)\, d\nu < \varepsilon.$$

*Here $\mathbf{1}_\xi$ denotes the indicator function of $\xi$.*

**Definition 2.** *Let $\boldsymbol{a} = (a_1, \ldots, a_d) \in \mathbb{R}^d$, with each $a_i \geq 1$. Let $T$ be a positive integer, and let $\xi \subset \mathrm{Gr}(d,n)$. Then we define $G(\boldsymbol{a}, \xi, T)$ to be the set of sublattices $\Lambda$ of $\mathbb{Z}^n$ of dimension $d$ with determinant at most $T$, such that*

$$\frac{\lambda_{i+1}(\Lambda)}{\lambda_i(\Lambda)} \geq a_i \text{ for all } i = 1, 2, \ldots, d,$$

*and $\Lambda_{\mathbb{R}} \in \xi$.*

The result of Schmidt that we intend to use is a combination of Theorems 3 and 5 in [9]:

**Theorem 2.** *Assuming $\xi \subset \mathrm{Gr}(d,n)$ is Jordan measurable, we have*

$$|G(\boldsymbol{a}, \xi, T)| \asymp \left( \prod_{i=1}^{d-1} a_i^{-i(d-i)} \right) \nu(\xi)\, T^n,$$

*where $f \asymp g$ means $f \ll g$ and $g \ll f$.*

Let $G(d,n,T)$ denote the set of all sublattices of $\mathbb{Z}^n$ of dimension $d$ with determinant at most $T$. Let $\mathbf{P} = \mathbf{P}_{d,n,T}$ denote the uniform probability distribution over $G(d,n,T)$.

**Corollary 1.** *For $t > 1$, we have*

$$\mathbf{P}\left( \max_{i \in [d]} \left\{ \frac{\lambda_{i+1}(\Lambda)}{\lambda_i(\Lambda)} \right\} \geq t \right) \ll (d-1)\, t^{-(d-1)}.$$

*Proof.* Following Aliev and Henk in [2], let

$$\boldsymbol{\delta}_i(t) := \Big(1, \ldots, 1, \underset{i}{t}, 1, \ldots, 1\Big)^\top \in \mathbb{R}^d.$$

Applying the union bound to Theorem 2, this probability is at most

$$\sum_{i=1}^{d-1} \frac{|G(\boldsymbol{\delta}_i(t), \mathrm{Gr}(d,n), T)|}{|G(\boldsymbol{\delta}_i(1), \mathrm{Gr}(d,n), T)|} \ll \sum_{i=1}^{d-1} t^{-i(d-i)} \leq (d-1)\, t^{-(d-1)}.$$

## 4 Typical Cramer's rule ratios

This section is devoted to showing that the largest absolute value of any entry of the matrix $\boldsymbol{A}_\sigma^{-1} \boldsymbol{A}_{\bar{\sigma}}$ is typically not too large, when the subspace $L := \ker \boldsymbol{A}$ is chosen uniformly at random from $\mathrm{Gr}(d,n)$. Note that the matrix $\boldsymbol{A}_\sigma^{-1} \boldsymbol{A}_{\bar{\sigma}}$ depends only on $L$ and $\sigma$. We remark that the entries of the matrix $\boldsymbol{A}_\sigma^{-1} \boldsymbol{A}_{\bar{\sigma}}$ are explicitly computed using Cramer's rule: for $i \in \sigma$ and $j \notin \sigma$, we have

$$\left(\boldsymbol{A}_\sigma^{-1} \boldsymbol{A}_j\right)_i = \frac{\det(\boldsymbol{A}_{\sigma - i + j})}{\det(\boldsymbol{A}_\sigma)}.$$

### 4.1 The real Grassmannian

For a general introduction to matrix groups and Grassmannians, we refer the reader to [3]. There is a right action of the orthogonal group $O(n)$ on $\mathrm{Gr}(d,n)$ defined as follows: if $\ker(\boldsymbol{A}) \in \mathrm{Gr}(d,n)$, where $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, then

$$(\ker(\boldsymbol{A})) \cdot \boldsymbol{U} = \ker(\boldsymbol{AU}). \tag{8}$$

This is well-defined, since if $\ker(\boldsymbol{A}) = \ker(\boldsymbol{A}')$ for some $\boldsymbol{A}' \in \mathbb{R}^{m\times n}$, then $\boldsymbol{A} = D\boldsymbol{A}'$ for some invertible $m \times m$ matrix $D$, and hence

$$\ker(\boldsymbol{AU}) = \ker(D\boldsymbol{A}'\boldsymbol{U}) = \ker(\boldsymbol{A}'\boldsymbol{U}).$$

Let $\mathrm{St}^{m\times n} := \{\boldsymbol{A} \in \mathbb{R}^{m\times n} : \mathrm{rank}(\boldsymbol{A}) = m\}$. Call this the *Stiefel manifold*. Again, there is a right action of $O(n)$ on $\mathrm{St}^{m\times n}$ which in this case is simply right multiplication:

$$\boldsymbol{A} \cdot \boldsymbol{U} = \boldsymbol{AU}.$$

The only thing to check here is that $\boldsymbol{AU}$ indeed lies in $\mathrm{St}^{m\times n}$, but this is indeed the case since

$$\boldsymbol{AU}(\boldsymbol{AU})^\top = \boldsymbol{AUU}^\top\boldsymbol{A}^\top = \boldsymbol{AA}^\top,$$

thus $\boldsymbol{A}$ and $\boldsymbol{AU}$ have the same Gram matrix $\boldsymbol{AA}^\top$, and a $m \times n$ matrix has full-rank if and only if its Gram matrix does.

The kernel map gives rise to a surjective map

$$\begin{aligned} \ker : \mathrm{St}^{m\times n} &\to \mathrm{Gr}(d,n) \\ \boldsymbol{A} &\mapsto \ker(\boldsymbol{A}) \end{aligned}$$

Thus, we see from (8) that the following statement holds:

**Proposition 2.** *The map* $\ker : \mathrm{St}^{m\times n} \to \mathrm{Gr}(d,n)$ *is equivariant with respect to the right actions of* $O(n)$ *on* $\mathrm{St}^{m\times n}$ *and* $\mathrm{Gr}(d,n)$*; that is,* $(\ker(\boldsymbol{A})) \cdot \boldsymbol{U} = \ker(\boldsymbol{A} \cdot \boldsymbol{U})$.

### 4.2 Probability spaces

Consider the probability space $(\mathbb{R}^{m\times n}, \mathscr{B}(\mathbb{R}^{m\times n}), \gamma)$ where $\mathscr{B}(\mathbb{R}^{m\times n})$ is the Borel $\sigma$-algebra, and the measure $\gamma$ is defined so that each $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ has iid $N(0,1)$ entries. In other words, $\gamma$ is the standard Gaussian probability measure on the $mn$-dimensional real vector space $\mathbb{R}^{m\times n}$ with mean zero and identity covariance matrix. By restricting to $\mathrm{St}^{m\times n}$, we get the probability space $(\mathrm{St}^{m\times n}, \mathscr{B}(\mathrm{St}^{m\times n}), \gamma)$. We can do this because $\mathbb{R}^{m\times n}\backslash\mathrm{St}^{m\times n}$ is an algebraic hypersurface in $\mathbb{R}^{m\times n}$, and therefore has measure zero with respect to $\gamma$. Let $\mathscr{B} := \mathscr{B}(\mathrm{St}^{m\times n})$.

The Grassmannian $\mathrm{Gr}(d,n)$ is endowed with the topology where $E \subseteq \mathrm{Gr}(d,n)$ is open iff $\ker^{-1}(E)$ is open in $\mathrm{St}^{m\times n}$. Let $\mathscr{G}$ denote the associated Borel $\sigma$-algebra. As before, we let $\nu : \mathscr{G} \to [0,1]$ denote the $O(n)$-invariant probability measure on $\mathrm{Gr}(d,n)$. This measure is characterized as follows:

**Proposition 3 ([7, Corollary 3.1.3]).** *The measure $\nu$ is the unique measure on $\mathrm{Gr}(d,n)$ satisfying*

$$\begin{aligned} \nu(E\cdot \boldsymbol{U}) &= \nu(E) \text{ for all } E\in\mathscr{G} \text{ and } \boldsymbol{U}\in O(n) \\ \nu(\mathrm{Gr}(d,n)) &= 1. \end{aligned} \tag{9}$$

The map $\ker : \mathrm{St}^{m\times n}\to \mathrm{Gr}(d,n)$ thus defines a map of probability spaces:

$$\ker : \left(\mathrm{St}^{m\times n},\mathscr{B},\gamma\right)\to\left(\mathrm{Gr}(d,n),\mathscr{G},\nu\right).$$

**Proposition 4.** *The measure $\nu$ is the pushforward measure of $\gamma$ under this map. That is, $\nu(E)=\gamma(\ker^{-1}(E))$ for each $E\in\mathscr{G}$.*

*Proof.* We establish the conditions of (9). By surjectivity, and the fact that $\gamma$ is a probability measure, we have

$$\gamma(\ker^{-1}(\mathrm{Gr}(d,n))) = \gamma\left(\mathrm{St}^{m\times n}\right) = 1.$$

It therefore remains to show $\gamma(\ker^{-1}(E\cdot\boldsymbol{U}))=\gamma(\ker^{-1}(E))$ for each $E\in\mathscr{G}$ and $\boldsymbol{U}\in O(n)$. By Proposition 2, we have

$$\ker^{-1}(E\cdot\boldsymbol{U}) = \ker^{-1}(E)\cdot\boldsymbol{U}. \tag{10}$$

Now, $\mathbb{R}^{m\times n}$ has the inner product $\langle \boldsymbol{A},\boldsymbol{B}\rangle = \mathrm{trace}\left(\boldsymbol{A}\boldsymbol{B}^{\top}\right)$. With respect to this inner product we may consider the subgroup $O(m\times n)$ of $\mathrm{GL}(\mathbb{R}^{m\times n})$ which is given by

$$O(m\times n) := \left\{\varphi\in\mathrm{GL}\left(\mathbb{R}^{m\times n}\right) : \langle\varphi(\boldsymbol{A}),\varphi(\boldsymbol{B})\rangle = \langle \boldsymbol{A},\boldsymbol{B}\rangle\right\}.$$

Observe that, for a fixed $\boldsymbol{U}\in O(n)$, the linear map $\varphi_{\boldsymbol{U}}\in\mathrm{GL}(\mathbb{R}^{m\times n})$ given by

$$\varphi_{\boldsymbol{U}}(\boldsymbol{A}) = \boldsymbol{A}\boldsymbol{U} \tag{11}$$

lies in $O(m\times n)$, since

$$\langle\varphi(\boldsymbol{A}),\varphi(\boldsymbol{B})\rangle = \mathrm{trace}\left(\boldsymbol{A}\boldsymbol{U}(\boldsymbol{B}\boldsymbol{U})^{\top}\right) = \mathrm{trace}\left(\boldsymbol{A}\boldsymbol{B}^{\top}\right) = \langle\boldsymbol{A},\boldsymbol{B}\rangle.$$

Now the probability measure $\gamma$ on $\mathbb{R}^{m\times n}$ is defined so that the coordinates $\boldsymbol{A}_{i,j}$ of a randomly chosen $\boldsymbol{A}\in\mathbb{R}^{m\times n}$ are iid $N(0,1)$ normally distributed. In particular this measure is invariant under isometry, in that for all $\mathcal{K}\in\mathscr{B}(\mathbb{R}^{m\times n})$ and $\varphi\in O(m\times n)$ ,we have

$$\gamma(\varphi(\mathcal{K})) = \gamma(\mathcal{K}). \tag{12}$$

The same is therefore true for the restricted probability measure $\gamma$ on $\mathrm{St}^{m\times n}$. It follows that if $\boldsymbol{U}\in O(n)$ and $E\in\mathscr{G}$, then, using (10), (11), and (12), we have

$$\gamma\left(\ker^{-1}(E\cdot\boldsymbol{U})\right) = \gamma\left(\ker^{-1}(E)\cdot\boldsymbol{U}\right) = \gamma\left(\varphi_{\boldsymbol{U}}\left(\ker^{-1}(E)\right)\right) = \gamma\left(\ker^{-1}(E)\right).$$

### 4.3 Cramer's rule ratios

Let $\sigma \subset [n]$ of size $m$, and define

$$\begin{aligned} \mathrm{St}_\sigma^{m\times n} &:= \left\{ \boldsymbol{A} \in \mathrm{St}^{m\times n} : \boldsymbol{A}_\sigma \text{ is nonsingular} \right\}. \\ \mathrm{Gr}\left(d,n\right)_\sigma &:= \left\{ \ker\left(\boldsymbol{A}\right) \in \mathrm{Gr}\left(d,n\right) : \boldsymbol{A}_\sigma \text{ is nonsingular} \right\}. \end{aligned}$$

Note that $\gamma\left(\mathrm{St}_\sigma^{m\times n}\right) = \nu\left(\mathrm{Gr}\left(n,d\right)_\sigma\right) = 1$. Also define, for $s > 1$, $i \in \sigma$, and $j \notin \sigma$,

$$\xi_{\sigma,i,j}\left(s\right) := \left\{ \ker\left(\boldsymbol{A}\right) \in \mathrm{Gr}\left(d,n\right)_\sigma : \left|\left(\boldsymbol{A}_\sigma^{-1}\boldsymbol{A}_j\right)_i\right| > s \right\}.$$

**Proposition 5.** *For $s > 1$ and $\sigma, i, j$ as above, we have*

$$\nu\left(\xi_{\sigma,i,j}\left(s\right)\right) = \frac{2}{\pi s} + \mathcal{O}\left(s^{-3}\right).$$

*Proof.* Let $\boldsymbol{A}$ be a random element of $\mathrm{St}_\sigma^{m\times n}$, and let $H$ denote the (random) hyperplane spanned by the columns of $\boldsymbol{A}_{\sigma-i}$, and let $\ell$ denote the line perpendicular to $H$. Let $\boldsymbol{u}_\ell$ denote the unit normal vector to $H$ whose first nonzero coordinate is positive. Thus,

$$\ell = \mathbb{R}\boldsymbol{u}_\ell = \left\{\lambda \boldsymbol{u}_\ell : \lambda \in \mathbb{R}\right\}.$$

Let $\alpha \in \{-1, +1\}$ denote the sign of the first nonzero entry of $\mathbf{e}_i^\top \boldsymbol{A}_\sigma^{-1}$. Then we can write

$$\boldsymbol{u}_\ell^\top = \frac{\alpha \mathbf{e}_i^\top \boldsymbol{A}_\sigma^{-1}}{\left\|\mathbf{e}_i^\top \boldsymbol{A}_\sigma^{-1}\right\|_2},$$

since for all $k \in \sigma - i$ we have

$$\alpha \mathbf{e}_i^\top \boldsymbol{A}_\sigma^{-1}\boldsymbol{A}_k = \alpha \mathbf{e}_i^\top \boldsymbol{A}_\sigma^{-1}\boldsymbol{A}_\sigma \mathbf{e}_k = 0,$$

and $\alpha \mathbf{e}_i^\top \boldsymbol{A}_\sigma^{-1}$ has first nonzero component positive by definition of $\alpha$.

Now let $k$ be any element of $[n]$ outside of $\sigma - i$. Since $\boldsymbol{u}_\ell$ depends only on $\boldsymbol{A}_{\sigma-i}$, and the entries of $\boldsymbol{A}$ are mutually independent, we have that $\boldsymbol{u}_\ell$ and $\boldsymbol{A}_k$ are independent random vectors. Now, for any fixed unit vector $\boldsymbol{v} \in \mathbb{S}^{n-1}$, as $\boldsymbol{A}_k$ has $N(0,1)$ iid entries, then the dot product $\boldsymbol{v}^\top \boldsymbol{A}_k$ also has distribution $N(0,1)$. Thus, for any fixed $t \in \mathbb{R}$, the random variable

$$\gamma\left(\boldsymbol{u}_\ell^\top \boldsymbol{A}_k \le t \mid \ell\right)$$

(i.e. the conditional probability in terms of the $\sigma$-algebra generated by $\ell$) is in fact constant. Evaluating at the line $\ell = \mathbb{R}\mathbf{e}_1$, for example, this constant is given by

$$\gamma\left(\boldsymbol{A}_{1,k} \le t\right).$$

This shows that the random quantity $\boldsymbol{u}_\ell^\top \boldsymbol{A}_k$ has distribution $N(0,1)$. We have

$$\left(\boldsymbol{A}_\sigma^{-1}\boldsymbol{A}_j\right)_i = \frac{\mathbf{e}_i^\top \boldsymbol{A}_\sigma^{-1}\boldsymbol{A}_j}{\mathbf{e}_i^\top \boldsymbol{A}_\sigma^{-1}\boldsymbol{A}_i} = \frac{\boldsymbol{u}_\ell^\top \boldsymbol{A}_j}{\boldsymbol{u}_\ell^\top \boldsymbol{A}_i}.$$

The independence of $\boldsymbol{u}_\ell^\top \boldsymbol{A}_i$ and $\boldsymbol{u}_\ell^\top \boldsymbol{A}_j$ imply that $\left(\boldsymbol{A}_\sigma^{-1}\boldsymbol{A}_j\right)_i$ has the Cauchy distribution, that is, the ratio of two iid $N(0,1)$ random variables. In particular, the cdf of $\left(\boldsymbol{A}_\sigma^{-1}\boldsymbol{A}_j\right)_i$ is given by

$$\gamma\left(\left(\boldsymbol{A}_\sigma^{-1}\boldsymbol{A}_j\right)_i \leq t\right) = \frac{1}{\pi}\arctan(t) + \frac{1}{2}.$$

See [6, p. 50] for more on the Cauchy distribution. Using the series expansion

$$\arctan(t) = \frac{\pi}{2} - \frac{1}{t} + \frac{1}{3t^3} - \frac{1}{5t^5} + \cdots,$$

we get

$$\gamma\left(\left(\boldsymbol{A}_\sigma^{-1}\boldsymbol{A}_j\right)_i \leq t\right) = 1 - \left(\frac{1}{\pi t} - \frac{1}{3\pi t^3} + \frac{1}{5\pi t^5} - \cdots\right).$$

Hence, using Proposition 4 and the fact $s > 1$, we conclude

$$\begin{aligned}
\nu\left(\xi_{\sigma,i,j}(s)\right) &= \gamma\left(\left|\left(\boldsymbol{A}_\sigma^{-1}\boldsymbol{A}_j\right)_i\right| > s\right) \\
&= 2\cdot\gamma\left(\left(\boldsymbol{A}_\sigma^{-1}\boldsymbol{A}_j\right)_i > s\right) \\
&= 2\left(1 - \gamma\left(\left(\boldsymbol{A}_\sigma^{-1}\boldsymbol{A}_j\right)_i \leq s\right)\right) \\
&= 2\left(\frac{1}{\pi s} - \frac{1}{3\pi s^3} + \frac{1}{5\pi s^5} - \cdots\right) \\
&= \frac{2}{\pi s} + \mathcal{O}\left(s^{-3}\right).
\end{aligned}$$

## 5 Proof of main result

In this final section we prove the main result of this paper, Theorem 1.

**Definition 3.** *Define the constant*

$$\tilde{\omega}_d := \frac{\omega_d^{1/d}}{d},$$

*where $\omega_d$ denotes the volume of the $d$-dimensional Euclidean ball. This constant $\tilde{\omega}_d$ is of the order $d^{-3/2}$.*

**Definition 4.** *Assume $\Lambda_{\mathbb{R}} = \ker(\boldsymbol{A})$. Given positive real numbers $s$ and $u$, we say $\Lambda$ is* $(\sigma, s, u)$-controlled *if $\sigma$ is a basis of $\boldsymbol{A}$ and:*

1. *The largest entry of $\boldsymbol{A}_\sigma^{-1}\boldsymbol{A}_{\bar{\sigma}}$ is at most $s$, and*
2. *The successive minima ratios of $\Lambda$ are not too large: we have*

$$\frac{\lambda_{i+1}(\Lambda)}{\lambda_i(\Lambda)} < \left(\tilde{\omega}_d u\right)^{2/(d-1)}$$

*for all $i = 1, 2, \ldots, d-1$.*

**Lemma 1 ([2, Proof of Lemma 5.2]).** *Suppose that*

$$\frac{\lambda_{i+1}(\Lambda)}{\lambda_i(\Lambda)} < \left(\tilde{\omega}_d u\right)^{2/(d-1)}$$

*for all* $i = 1, 2, \ldots, d-1$. *Then*

$$\mu < u\left(\Delta(\Lambda)\right)^{1/d}.$$

**Lemma 2.** *If* $\sigma$ *is a basis of* $\boldsymbol{A}$ *and* $\Lambda$ *is* $(\sigma, s, u)$*-controlled, then for all* $\boldsymbol{x}^* \in \mathcal{S}_\sigma$ *we have*

$$\operatorname{dist}(\Lambda, \sigma, \boldsymbol{x}^*) \leq 2n^{3/2} su\left(\Delta(\Lambda)\right)^{1/d}.$$

*Proof.* Let $\boldsymbol{b} = \boldsymbol{A}\boldsymbol{x}^*$, let $B = B_2^n \cap \Lambda_{\mathbb{R}}$, and let $\mu$ denote the covering radius of $B$ with respect to $\Lambda$. Define the vector $\boldsymbol{v} \in \mathbb{R}^n$ so that:

$$\begin{aligned} \boldsymbol{v}_j &= \mu \boldsymbol{w}_j \text{ for all } j \in \bar{\sigma} \\ \boldsymbol{A}\boldsymbol{v} &= \boldsymbol{b}. \end{aligned}$$

We show that the scaled, translated ball $\mu B + \boldsymbol{v}$ is contained in $\mathcal{P}(\boldsymbol{A}, \boldsymbol{b})$. Since $B \subseteq \Lambda_{\mathbb{R}}$, we have that each $\boldsymbol{x} \in \mu B + \boldsymbol{v}$ satisfies $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$. For each $j \in [n]$, let $\boldsymbol{x}^{(j)}$ be the unique point in $\mu B + \boldsymbol{v}$ such that $\boldsymbol{x}_j^{(j)}$ is minimized. If $j \in \bar{\sigma}$, then

$$\boldsymbol{x}_j^{(j)} = \mu(-\boldsymbol{w}_j) + \boldsymbol{v}_j = \mu(-\boldsymbol{w}_j) + \mu \boldsymbol{w}_j = 0.$$

If $j \in \sigma$, then since $\boldsymbol{x}^* \in \mathcal{S}_\sigma$ we have

$$\begin{aligned} \boldsymbol{x}_j^{(j)} &= \mu(-\boldsymbol{w}_j) + \boldsymbol{v}_j \\ &= \mu(-\boldsymbol{w}_j) + \left(\boldsymbol{A}_\sigma^{-1}\boldsymbol{b} - \boldsymbol{A}_\sigma^{-1}\boldsymbol{A}_{\bar{\sigma}}\boldsymbol{w}_{\bar{\sigma}}\right)_j \\ &\geq \mu(-\boldsymbol{w}_j) + \mu \boldsymbol{w}_j \\ &= 0. \end{aligned}$$

This concludes the proof that $\mu B + \boldsymbol{v} \subseteq \mathcal{P}(\boldsymbol{A}, \boldsymbol{b})$.

Now, since $\mu$ is the covering radius of $B$ with respect to $\Lambda$, there exists $\boldsymbol{z}^* \in \mathbb{Z}^n \cap (\mu B + \boldsymbol{v})$ such that

$$\|\boldsymbol{x}^* - \boldsymbol{z}^*\|_2 \leq \|\boldsymbol{x}^* - \boldsymbol{v}\|_2 + \|\boldsymbol{v} - \boldsymbol{z}^*\|_2 \leq \mu \|\tilde{\boldsymbol{w}}\|_2 + \mu. \tag{13}$$

where we define $\tilde{\boldsymbol{w}} := (\boldsymbol{v} - \boldsymbol{x}^*)/\mu$. That is, $\tilde{\boldsymbol{w}}$ satisfies

$$\begin{aligned} \boldsymbol{A}\tilde{\boldsymbol{w}} &= \boldsymbol{0} \\ \tilde{\boldsymbol{w}} &= \boldsymbol{w}_j \text{ for all } j \in \bar{\sigma}. \end{aligned}$$

Observe that

$$\tilde{\boldsymbol{w}}_\sigma = -\boldsymbol{A}_\sigma^{-1}\boldsymbol{A}_{\bar{\sigma}}\tilde{\boldsymbol{w}}_{\bar{\sigma}}.$$

Using the fact $\boldsymbol{w} \in [0,1]^n$, we therefore have

$$\begin{aligned}\|\tilde{\boldsymbol{w}}\|_2^2 &= \|\tilde{\boldsymbol{w}}_\sigma\|_2^2 + \|\tilde{\boldsymbol{w}}_{\bar{\sigma}}\|_2^2 \\ &= \left\|\boldsymbol{A}_\sigma^{-1}\boldsymbol{A}_{\bar{\sigma}}\tilde{\boldsymbol{w}}_{\bar{\sigma}}\right\|_2^2 + \|\tilde{\boldsymbol{w}}_{\bar{\sigma}}\|_2^2 \\ &\leq m\left\|\boldsymbol{A}_\sigma^{-1}\boldsymbol{A}_{\bar{\sigma}}\right\|_\infty^2 \|\tilde{\boldsymbol{w}}_{\bar{\sigma}}\|_1^2 + \|\tilde{\boldsymbol{w}}_{\bar{\sigma}}\|_2^2 \\ &\leq \left(ms^2+1\right)d^2.\end{aligned}$$

Thus we conclude

$$\begin{aligned}\|\boldsymbol{x}^* - \boldsymbol{z}^*\|_2 &\leq \mu\left(\|\tilde{\boldsymbol{w}}\|_2 + 1\right) \\ &\leq u\Delta^{1/d}\left(\sqrt{\left(ms^2+1\right)d^2}+1\right) \\ &\leq 2n^{3/2}su\Delta^{1/d}.\end{aligned}$$

*Proof (Proof of Theorem 1).* Let $\Lambda$ be a uniformly chosen lattice from $G(d,n,T)$. Let $t > 1$, and let $s := t^{2/3}/(2n^{3/2})$ and $u := t^{1/3}$, so that $t = 2n^{3/2}su$ as in Lemma 2. We have

$$\begin{aligned}&\mathbf{P}\left(\operatorname{dist}(\Lambda) > t\left(\Delta(\Lambda)\right)^{1/d}\right) \\ &\leq \sum_\sigma \mathbf{P}\left(\sigma \text{ basis of } \boldsymbol{A},\ \operatorname{dist}(\Lambda,\sigma,\boldsymbol{x}^*) > t\left(\Delta(\Lambda)\right)^{1/d} \text{ for some } \boldsymbol{x}^* \in \mathcal{S}_\sigma\right) \\ &\leq \sum_\sigma \mathbf{P}\left(\sigma \text{ basis of } \boldsymbol{A},\ \Lambda \text{ is not } (\sigma,s,u)\text{-controlled}\right)\end{aligned}$$

where the sums are over all subsets $\sigma \subseteq [n]$ of size $m$. It therefore suffices to show, for each such $\sigma$,

$$\mathbf{P}\left(\sigma \text{ basis of } \boldsymbol{A},\ \Lambda \text{ is not } (\sigma,s,u)\text{-controlled}\right) \ll t^{-2/3}.$$

By definition, this probability is at most

$$\mathbf{P}\left(\max_{i\in[d]}\left\{\frac{\lambda_{i+1}(\Lambda)}{\lambda_i(\Lambda)}\right\} \geq \left(\tilde{\omega}_d u\right)^{2/(d-1)}\right) + \sum_{\substack{i\in\sigma \\ j\notin\sigma}} \mathbf{P}\left(\sigma \text{ basis of } \boldsymbol{A},\ \left(\boldsymbol{A}_\sigma^{-1}\boldsymbol{A}_j\right)_i \geq s\right). \tag{14}$$

By Theorem 2, we have

$$\mathbf{P}\left(\sigma \text{ basis of } \boldsymbol{A},\ \left(\boldsymbol{A}_\sigma^{-1}\boldsymbol{A}_j\right)_i \geq s\right) = \frac{\left|G\left(\mathbf{1},\xi_{\sigma,i,j}(s),T\right)\right|}{\left|G\left(\mathbf{1},\operatorname{Gr}(d,n),T\right)\right|} \asymp \nu\left(\xi_{\sigma,i,j}(s)\right).$$

Hence, applying Corollary 1 and Proposition 5, for $T$ sufficiently large, we may estimate up to constants the quantity (14) by

$$u^{-2} + s^{-1} \ll t^{-2/3}.$$

## References


1. I. Aliev, M. Henk, and T. Oertel. Distances to lattice points in knapsack polyhedra. *Math. Program.*, 182(1-2, Ser. A):175–198, 2020.
2. Iskander Aliev and Martin Henk. Feasibility of integer knapsacks. *SIAM Journal on Optimization*, 20(6):2978–2993, 2010.
3. A. Baker. *Matrix Groups: An Introduction to Lie Group Theory*. Springer Undergraduate Mathematics Series. Springer London, 2003.
4. W. Cook, A. M. H. Gerards, A. Schrijver, and É. Tardos. Sensitivity theorems in integer linear programming. *Mathematical Programming*, 34(3):251–264, Apr 1986.
5. Friedrich Eisenbrand and Robert Weismantel. Proximity results and faster algorithms for integer programming using the Steinitz lemma. *ACM Trans. Algorithms*, 16(1), November 2019.
6. V. Feller and W. Feller. *An Introduction to Probability Theory and Its Applications, Volume 1*. A Wiley publication in mathematical statistics. Wiley, 1968.
7. S.G. Krantz and H.R. Parks. *Geometric Integration Theory*. Cornerstones. Birkhäuser Boston, 2008.
8. Timm Oertel, Joseph Paat, and Robert Weismantel. The distributions of functions related to parametric integer optimization. *SIAM Journal on Applied Algebra and Geometry*, 4(3):422–440, 2020.
9. Wolfgang M Schmidt. The distribution of sublattices of $\mathbf{Z}^m$. *Monatshefte für Mathematik*, 125(1):37–81, 1998.