Mathematical Programming (2023) 197:1201–1219
https://doi.org/10.1007/s10107-022-01786-8

FULL LENGTH PAPER

Series B

Proximity bounds for random integer programs

Marcel Celaya[1] · **Martin Henk**[2]

Received: 31 May 2021 / Accepted: 30 December 2021 / Published online: 18 March 2022

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 $\boldsymbol{A} \in \mathbb{Z}^{m\times n}$ is of rank m, $\boldsymbol{b} \in \mathbb{Z}^m$ and $\boldsymbol{c} \in \mathbb{Z}^n$. In particular, we seek bounds for proximity in terms of the parameter $\Delta(\boldsymbol{A})$, which is the square root of the determinant of the Gram matrix $\boldsymbol{A}\boldsymbol{A}^\top$ of $\boldsymbol{A}$. We prove that, up to constants depending on n and m, the proximity is "generally" bounded by $\Delta(\boldsymbol{A})^{1/(n-m)}$, which is significantly better than the best deterministic bounds which are, again up to dimension constants, linear in $\Delta(\boldsymbol{A})$.

Mathematics Subject Classification 11H06 · 52C07 · 52C17 · 90C10

1 Introduction

Given a linear program of the form

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

where $\boldsymbol{A}$ is a full-row-rank $m \times n$ integral matrix, $\boldsymbol{b} \in \mathbb{Z}^m$, and $\boldsymbol{c} \in \mathbb{Z}^n$, we seek to understand how far away an optimal vertex $\boldsymbol{x}^*$ of the feasible region can be to a nearby feasible integer solution $\boldsymbol{z}^*$, assuming the feasible region has at least one integral point. Typically it is further required that $\boldsymbol{z}^*$ is itself optimal; we do not impose this requirement in this manuscript. We refer to the smallest possible distance between $\boldsymbol{x}^*$

✉ Marcel Celaya
marcel.celaya@ifor.math.ethz.ch

Martin Henk
henk@math.tu-berlin.de

[1] Department of Mathematics, Institute for Operations Research, ETH Zurich, Rämistrasse 101, 8092 Zurich, Switzerland

[2] Institut für Mathematik, Technische Universität Berlin, Sekr. MA4-1, Straße des 17 Juni 136, 10623 Berlin, Germany

and a feasible integral solution z^* as the *proximity* of (1). This distance is measured in terms of some given norm, for example the $\|\cdot\|_1$ or $\|\cdot\|_\infty$ norms; in this paper we state our results in terms of the Euclidean norm $\|\cdot\|_2$.

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 $\boldsymbol{A}$. Note that this parameter is within a factor of $\binom{n}{m}$ of $\Delta(\boldsymbol{A}) := \sqrt{\det(\boldsymbol{A}\boldsymbol{A}^\top)}$. Finding such bounds is a well-studied problem which goes back to the classic Cook et al. result [7] bounding the proximity of the dual of (1). See, for instance, the recent works of Eisenbrand and Weismantel [8] and of Aliev et al. [2] and the references therein.

In this manuscript, we would like to understand the worst-possible proximity, which we denote by $\mathrm{dist}(\boldsymbol{A})$, 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 $\boldsymbol{A}$ uniformly at random subject to the condition that the greatest common divisor of the maximal minors of $\boldsymbol{A}$ is 1, and that $\Delta(\boldsymbol{A})$ is at most some sufficiently large (with respect to m and n) integer T. This is a natural model to study from a geometric point of view, since $\Delta(\boldsymbol{A})$ 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 [1], in their investigation of diagonal Frobenius numbers.

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

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

Here $\ll$ means less than, up to constants which only depend on n and m. In particular, this shows that proximity generally depends only on $\Delta^{1/(n-m)}$ in our random setting, for "almost all" choices of $\boldsymbol{b}$ in a certain precise sense. This is significantly better than the linear dependency on Δ_m in the deterministic case, that is known to be tight [2, Theorem 1].

1.1 Related work

A similar result, with a slightly different random model, was obtained in [2] the so-called knapsack scenario, where $m = 1$. In this work, a fixed integer T is given, and the matrix $\boldsymbol{A}$ is a row vector chosen uniformly at random from $\{1, 2, \ldots, T\}^n$ such that the greatest common divisor of the entries equals 1. A special case of [2, Theorem 2] states

$$\mathbf{P}\left(\mathrm{dist}(\boldsymbol{A}) > t\,\|\boldsymbol{A}\|_\infty^{2/n}\right) \ll t^{-1},$$

where $\mathrm{dist}(\boldsymbol{A})$ measures distance using the $\|\cdot\|_\infty$ norm.

The recent work of Oertel et al. [14] considers a random model that allows $\boldsymbol{b}$ to vary but keeps $\boldsymbol{A}$ fixed. More precisely, for a given positive integer t, the vector $\boldsymbol{b}$ is chosen

uniformly at random from $\{-T, \ldots, T\}^m$ such that $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \geq \boldsymbol{0}$ is integer-feasible. The result in [14, Corollary 1.3] states that

$$\operatorname{dist}(\boldsymbol{A}) \leq (m+1)\left(\sqrt{m}\,\|\boldsymbol{A}\|_\infty\right)^m$$

with probability approaching 1 as $T \to \infty$. Here again $\operatorname{dist}(\boldsymbol{A})$ measures distance using the $\|\cdot\|_\infty$ norm. Note that this bound does not depend on n.

Finally, we mention the very recent work of Borst et al. [5] which investigates the *integrality gap* of integer programs of the form

$$\begin{aligned} \max\ \boldsymbol{c}^\top \boldsymbol{x}\ :\quad \boldsymbol{A}\boldsymbol{x} &\leq \boldsymbol{b} \\ \boldsymbol{0} \leq \boldsymbol{x} &\leq \boldsymbol{1} \\ \boldsymbol{x} &\in \mathbb{Z}^n, \end{aligned} \tag{3}$$

with $\boldsymbol{A}$ and $\boldsymbol{c}$ having independent, Gaussian $N(0, 1)$ entries. This quantity measures the difference between the optimal value of (3) and that of its linear relaxation. Their result is that the integrality gap is bounded from above by $\operatorname{poly}(m)\,(\log n)^2/n$ with probability at least $1 - n^{-7} - 2^{-\operatorname{poly}(m)}$, subject to certain conditions on $\boldsymbol{b}$. See [5] and the references therein for a history of this problem.

1.2 Outline of proof

The proof of our result combines ideas of [1,2] using facts from the geometry of numbers, some results of Schmidt from [15] on random sublattices of $\mathbb{Z}^n$ of fixed dimension, and computations of the measure of certain distinguished regions of the real Grassmannian $\operatorname{Gr}(d, n)$ of d-dimensional subspaces of $\mathbb{R}^n$, where $d = n - m$. For us the crucial parameters from the geometry of numbers that we need are the covering radius μ, as well as the successive minima $\lambda_1, \ldots, \lambda_d$ of $\ker \boldsymbol{A} \cap B_2^n$ with respect to the lattice $\ker \boldsymbol{A} \cap \mathbb{Z}^n$, where B_2^n denotes the unit-radius Euclidean ball in $\mathbb{R}^n$. Further details on these parameters can be found in Sect. 3.

The restrictions imposed by the definition of $\operatorname{dist}^*(\boldsymbol{A})$ on the right hand side $\boldsymbol{b}$ ensure that, given a vertex $\boldsymbol{x}^*$ of the feasible region of (1), one can always find a feasible integral solution $\boldsymbol{z}^*$ such that

$$\left\|\boldsymbol{x}^* - \boldsymbol{z}^*\right\|_2 \leq \mu\left(\left\|\boldsymbol{A}_\sigma^{-1}\boldsymbol{A}\right\|_1 + 1\right),$$

where $\boldsymbol{x}^*$ has support contained in $\sigma \subseteq [n]$ and $\boldsymbol{A}_\sigma$ denotes the square submatrix of $\boldsymbol{A}$ whose columns are indexed by σ. This restriction on $\boldsymbol{b}$ amounts to picking $\boldsymbol{b}$ sufficiently deep inside the cone spanned by the columns of $\boldsymbol{A}_\sigma$, or choosing $\boldsymbol{b}$ from a reduced cone in the sense of Gomory [11, p. 261]. A uniform upper bound on all ratios $\lambda_{i+1}/\lambda_i,\ i = 1, 2, \ldots, d-1$ implies an upper bound on μ, see Lemma 2. Meanwhile, Sect. 4 shows that the measure in $\operatorname{Gr}(d, n)$ of those subspaces $\ker \boldsymbol{A} \in \operatorname{Gr}(d, n)$ such that any given entry of $\boldsymbol{A}_\sigma^{-1}\boldsymbol{A}$ exceeds in absolute value some fixed parameter $s > 0$ is a function of the order s^{-1}. Theorem 2, itself a straightforward corollary of results

of [15], combines these two pieces together: a random lattice of the form $\ker \boldsymbol{A} \cap \mathbb{Z}^n$ is unlikely to have any ratio λ_{i+1}/λ_i, nor any entry of $\boldsymbol{A}_\sigma^{-1}\boldsymbol{A}$, exceedingly large. The details of this are carried out in Sect. 5.

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. When we allow $\boldsymbol{b}$ to lie close the the boundary of the cone spanned by the columns of $\boldsymbol{A}_\sigma$, our arguments no longer apply.

Remark 1 (*Changes from proceedings version*) The following changes have been made since the proceedings version of this manuscript [6]. In Sect. 3 we clarified and expanded upon the geometry of numbers theory that is used in this paper. In Sect. 4 we gave a proof of the claim that a particular subset of $\operatorname{Gr}(d, n)$ is Jordan measurable. Some minor typos have also been corrected, and some minor changes have been made to the introduction.

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 σ. 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 σ. 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 σ is a *basis* of $\boldsymbol{A}$. We denote the complement of σ 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 Λ. We say $\sigma \subseteq [n]$ is a *coordinate basis* of Λ or $\Lambda_\mathbb{R}$ if the coordinate projection map

$$\begin{aligned} \Lambda_\mathbb{R} &\to \mathbb{R}^\sigma \\ \boldsymbol{x} &\mapsto \boldsymbol{x}_\sigma \end{aligned}$$

is an isomorphism. This is equivalent to saying that σ is a basis of $\boldsymbol{A}$ for any full-row-rank matrix $\boldsymbol{A}$ such that $\ker(\boldsymbol{A}) = \Lambda_\mathbb{R}$. Finally, we denote the group of $n \times n$ orthogonal real matrices by $O(n)$. This notation presents a conflict with "big-O" asymptotic notation, so we write $\mathcal{O}(n)$ for the latter.

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

Let $\boldsymbol{A} \in \mathbb{Z}^{m\times n}$ be a full-row-rank matrix. For a basis σ of $\boldsymbol{A}$, we define the semigroup

$$\mathcal{S}_\sigma(\boldsymbol{A}) := \left\{ \boldsymbol{x} \geq \boldsymbol{0} : \boldsymbol{x}_{\bar{\sigma}} = \boldsymbol{0},\ \boldsymbol{x}_\sigma = \boldsymbol{A}_\sigma^{-1}\boldsymbol{A}\mathbf{g} \text{ for some } \mathbf{g} \in \mathbb{Z}^n_{\geq 0} \right\}. \tag{4}$$

For a vector $\boldsymbol{b} \in \mathbb{Z}^m$, we define the polyhedron

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

The idea behind these definitions is that if $\boldsymbol{x}^* \in \mathcal{S}_\sigma(\boldsymbol{A})$, then $\boldsymbol{b} := \boldsymbol{A}\boldsymbol{x}^*$ is an integral vector, $\mathcal{P}(\boldsymbol{A}, \boldsymbol{b})$ is a polyhedron containing at least one integral point, and $\boldsymbol{x}^*$ is the vertex of $\mathcal{P}(\boldsymbol{A}, \boldsymbol{b})$ associated to the basis σ. Now given a basis σ of $\boldsymbol{A}$ and $\boldsymbol{x}^* \in \mathcal{S}_\sigma(\boldsymbol{A})$, we define the distance

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

where $\boldsymbol{b} := \boldsymbol{A}\boldsymbol{x}^*$. We then define the worst-case distance over all choices of bases σ of $\boldsymbol{A}$ and elements $\boldsymbol{x}^* \in \mathcal{S}_\sigma(\boldsymbol{A})$ as

$$\operatorname{dist}(\boldsymbol{A}) := \max_{\sigma} \max_{\boldsymbol{x}^*} \operatorname{dist}\left(\boldsymbol{A}, \sigma, \boldsymbol{x}^*\right). \tag{5}$$

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 $m \times m$ integral 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 $\operatorname{dist}(\boldsymbol{A})$ which we now state.

2.3 Definition of dist(Λ)

Suppose instead we start with a d-dimensional sublattice Λ of $\mathbb{Z}^n$. Suppose σ is a coordinate basis of Λ. Then we may define the semigroup

$$\mathcal{S}_\sigma(\Lambda) := \left\{ \boldsymbol{x} \geq \boldsymbol{0} : \boldsymbol{x}_{\bar{\sigma}} = \boldsymbol{0},\ \boldsymbol{x} \in \Lambda_{\mathbb{R}} + \mathbf{g} \text{ for some } \mathbf{g} \in \mathbb{Z}^n_{\geq 0} \right\}. \tag{6}$$

For $\boldsymbol{x}^* \in \mathcal{S}_\sigma(\Lambda)$, define the distance

$$\operatorname{dist}\left(\Lambda, \sigma, \boldsymbol{x}^*\right) := \max_{\mathbf{g} \in (\Lambda_{\mathbb{R}} + \boldsymbol{x}^*) \cap \mathbb{Z}^n} \min_{\boldsymbol{z}^* \in (\Lambda + \mathbf{g}) \cap \mathbb{R}^n_{\geq 0}} \left\| \boldsymbol{x}^* - \boldsymbol{z}^* \right\|_2. \tag{7}$$

The extra maximum accounts for the fact that, if Λ is not primitive, then there are multiple ways to embed Λ into $\Lambda_{\mathbb{R}} + \boldsymbol{x}^*$ as an integral translate of Λ. Finally, define the worst case distance

$$\operatorname{dist}(\Lambda) := \max_{\sigma} \max_{\boldsymbol{x}^*} \operatorname{dist}\left(\Lambda, \sigma, \boldsymbol{x}^*\right), \tag{8}$$

where the maximum is taken over all coordinate bases of Λ and elements $\boldsymbol{x}^* \in \mathcal{S}_\sigma(\Lambda)$.

We now explain the relationship between definitions (5) and (8). First note that if $\boldsymbol{A}$ is any integral matrix such that $\Lambda_{\mathbb{R}} = \ker(\boldsymbol{A})$, then the two definitions (4) and (6) of $\mathcal{S}_\sigma(\boldsymbol{A})$ and $\mathcal{S}_\sigma(\Lambda)$ coincide. Moreover, if Λ is a *primitive* lattice, that is, if

$\Lambda = \Lambda_{\mathbb{R}} \cap \mathbb{Z}^n$, then we have

$$\operatorname{dist}\left(\Lambda, \sigma, \boldsymbol{x}^*\right) = \operatorname{dist}\left(\boldsymbol{A}, \sigma, \boldsymbol{x}^*\right)$$

and therefore

$$\operatorname{dist}(\Lambda) = \operatorname{dist}(\boldsymbol{A}).$$

Definition (8) also makes sense when Λ is non-primitive, however, and it is immediate from the definitions that in general,

$$\operatorname{dist}(\Lambda) \geq \operatorname{dist}\left(\Lambda_{\mathbb{R}} \cap \mathbb{Z}^n\right).$$

The key advantage of definition (8) is that there are only finitely many d-dimensional sublattices of $\mathbb{Z}^n$ whose determinant 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(Λ)

We next consider a modification of $\operatorname{dist}(\Lambda)$. Choose any full-row-rank matrix $\boldsymbol{A}$ such that $\ker(\boldsymbol{A}) = \Lambda_{\mathbb{R}}$, the particular choice of $\boldsymbol{A}$ is not important. Let $B_2^n \subset \mathbb{R}^n$ denote the n-dimensional Euclidean ball of radius 1.

Define the vector $\boldsymbol{w} = \boldsymbol{w}(\Lambda_{\mathbb{R}}) \in \mathbb{R}^n$ so that, for each $i \in [n]$,

$$\boldsymbol{w}_i := \max_{\boldsymbol{x} \in B_2^n \cap \Lambda_{\mathbb{R}}} \boldsymbol{x}_i.$$

Denote by $\mu = \mu(\Lambda)$ the covering radius of B_2^n with respect to Λ. That is,

$$\mu := \inf\left\{t > 0 : \Lambda + t B_2^n \text{ contains } \Lambda_{\mathbb{R}}\right\}. \tag{9}$$

For more information on the covering radius we refer to Sect. 3. If σ is a basis of $\boldsymbol{A}$ then define the following subsemigroup of $\mathcal{S}_\sigma(\Lambda)$:

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

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

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

Now set

$$\operatorname{dist}^*(\Lambda) := \max_{\sigma} \max_{\boldsymbol{x}^*} \operatorname{dist}\left(\Lambda, \sigma, \boldsymbol{x}^*\right), \tag{10}$$

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

2.5 Main result

We are now ready to state the main theorem.

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

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

What we would like to do is translate this statement into a statement about integer programs, and in particular derive inequality (2). For this we use a known result on the ratio between primitive sublattices and all sublattices with a fixed determinant upper bound, a consequence of Theorems 1 and 2 in [15]:

Lemma 1 *Suppose there are exactly $N(d, n, T)$ d-dimensional sublattices of $\mathbb{Z}^n$ with determinant at most T, of which exactly $P(d, n, T)$ are primitive. Then*

$$\lim_{T\to\infty} \frac{P(d,n,T)}{N(d,n,T)} = \frac{1}{\zeta(d+1)\cdots\zeta(n)},$$

where $\zeta(\cdot)$ denotes the Riemann zeta function.

Recall from the introduction our probability model. We start with a sufficiently large integer T relative to m and n, and consider the set of all $m \times n$ integral matrices $\boldsymbol{A}$ such that the greatest common divisor of all maximal minors of $\boldsymbol{A}$ equals 1, and that $\Delta(\boldsymbol{A}) \le T$. The group of $m \times m$ unimodular matrices acts on this set of matrices by multiplication on the left, and there are finitely many orbits of this action. We consider the uniform distribution on these orbits. We define

$$\operatorname{dist}^*(\boldsymbol{A}) := \operatorname{dist}^*\left(\ker(\boldsymbol{A}) \cap \mathbb{Z}^n\right).$$

Note that this definition depends not on $\boldsymbol{A}$ but only on the orbit of $\boldsymbol{A}$. The greatest common divisor condition ensures that $\Delta(\boldsymbol{A})$ equals the determinant of the lattice $\ker(\boldsymbol{A}) \cap \mathbb{Z}^n$. Recall we set $d := n - m$. We derive the next corollary by combining Theorem 1, Lemma 1, and the simple conditional probability inequality $\mathbf{P}(E \mid F) \le \mathbf{P}(E)/\mathbf{P}(F)$, where E is the event that $\operatorname{dist}^*(\Lambda) > t\,(\Delta(\Lambda))^{1/d}$ and F is the event that Λ is primitive.

Corollary 1 *For $T \gg 1$, choose $\boldsymbol{A}$ randomly as above, with determinant at most T. Then for all $t > 1$,*

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

We remark that the question of deriving the constants in this bound remains unexplored.

3 Geometry of Numbers and a theorem of Schmidt

Next we state some basic functionals and tools from Geometry of Numbers as well as a theorem of Schmidt which are fundamental for the proof of our results. An excellent reference for the Geometry of Numbers tools is Gruber's book [12, Chapters 21–23]. We start with Minkowski's successive minima. Given a d-dimensional lattice $\Lambda \subset \mathbb{R}^d$, the ith successive minimum $\lambda_i(\Lambda)$, $i \in \{1, \ldots, d\}$, is defined as

$$\lambda_i(\Lambda) := \min\{\lambda > 0 : \dim(\lambda\, B_2^d \cap \Lambda) \geq i\}.$$

In other words, $\lambda_i(\Lambda)$ is the smallest dilation factor λ such that the Euclidean ball of radius λ contains at least i linearly independent lattice points of Λ. Observe that

$$\lambda_1(\Lambda) \leq \lambda_2(\Lambda) \leq \cdots \leq \lambda_d(\Lambda).$$

Minkowski introduced these successive minima not only for a ball but for any convex body symmetric to the origin, but here we just need them for the ball. In this particular setting, Minkowski's so called second theorem on successive minima reads as follows

$$\lambda_1(\Lambda) \cdots \lambda_d(\Lambda)\, \omega_d \leq 2^d \det \Lambda, \tag{11}$$

where ω_d is the d-dimensional volume of the ball B_2^d. Inequality (11) is for $d > 1$ actually a strict inequality and one can improve on the factor 2^d on the right hand side, but for our purposes it is enough to use (11). The other functional we need from Geometry of Numbers is the already introduced covering radius $\mu(\Lambda)$ (see (9)) which may also be defined as

$$\mu(\Lambda) = \min\{\mu > 0 : (\boldsymbol{y} + \mu\, B_2^d) \cap \Lambda \neq \emptyset \text{ for all } \boldsymbol{y} \in \mathbb{R}^d\}.$$

The so called Jarnik's inequalities show that the covering radius is essentially of the size of the last succesive minimum

$$\frac{1}{2}\lambda_d(\Lambda) \leq \mu(\Lambda) \leq \frac{1}{2}\left(\lambda_1(\Lambda) + \cdots + \lambda_d(\Lambda)\right). \tag{12}$$

Now, in general the successive minima can take any arbitrary values, even for sublattices of $\mathbb{Z}^d$. A fundamental result of Schmidt [15] states, however, that for a "typical" primitive sublattice of $\mathbb{Z}^d$ the ratios $\lambda_{i+1}(\Lambda)/\lambda_i(\Lambda)$ are not "too" large. So one may expect that all the successive minima are more or less of the same size, which then allows us to give a "good" bound on $\mu(\Lambda)$ via (11) and (12). But first we need a few more definitions in order to state Schmidt's result.

We continue with our assumption that $d = n - m$. Let $\mathrm{Gr}\,(d, n)$ denote the set of d-dimensional subspaces of $\mathbb{R}^n$. Let ν denote the unique $O(n)$-invariant probability measure on the real Grassmannian $\mathrm{Gr}\,(d, n)$ (see, e.g., [3, Section 3.3]).

Definition 1 ([15, p. 40]) A subset $\xi \subset \mathrm{Gr}\,(d, n)$ 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 ξ.

In the next definiton we define the set $G\,(\boldsymbol{a}, \xi, T)$ of lattices we are interested in: they are sublattices of $\mathbb{Z}^d$ of determinant at most T, their span $\Lambda_\mathbb{R}$ is contained in a given subset $\xi \subseteq \mathrm{Gr}\,(d, n)$ and the ratios $\lambda_{i+1}(\Lambda)/\lambda_i(\Lambda)$ are at least as large as the ith entry of the given vector $\boldsymbol{a}$. More formally,

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 Λ of $\mathbb{Z}^n$ of dimension d with determinant at most T, such that

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

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

The result of Schmidt that we intend to use is a combination of Theorems 3 and 5 in [15]:

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 $g \ll f$.

Roughly speaking, the amount of lattices having large successive minima ratios is small. In order to formalize this, 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 2 *For $t > 1$, we have*

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

Proof Following Aliev and Henk [1], let

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

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

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

□

Finally we present the already mentioned upper bound on $\mu(\lambda)$ provided we know that $\lambda_{i+1}(\Lambda)/\lambda_i(\Lambda)$ is bounded. The argument is implicitly contained in the proof of Lemma 5.1 in [1].

Lemma 2 *Let $\Lambda \subset \mathbb{R}^d$ be a lattice, and let $u > 0$ such that for $1 \leq i \leq d-1$*

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

Then

$$\mu(\Lambda) \leq u\, (\det \Lambda)^{\frac{1}{d}}.$$

Proof For abbreviation we set $r := \left(u\omega_d^{1/d}/d\right)^{\frac{2}{d-1}}$. Due to our assumption we get a lower bound on all successive minima $\lambda_i(\Lambda)$, $i = 1, \ldots, d-1$, in terms of the last successive minimum

$$\lambda_d(\lambda) \leq r^{d-i} \lambda_i(\Lambda).$$

Combined with Minkowski's inequality (11) we obtain

$$\lambda_d(\Lambda)^d r^{-d(d-1)/2} \leq \lambda_1(\Lambda) \cdots \lambda_d(\Lambda) \leq \frac{2^d}{\omega_d} \det \Lambda.$$

Hence,

$$\lambda_d(\Lambda) \leq \left(u \frac{\omega_d^{1/d}}{d} \right) \frac{2}{\omega_d^{1/d}} \det \Lambda^{\frac{1}{d}} = \frac{2}{d} u (\det \Lambda)^{\frac{1}{d}},$$

and Jarnik's inequality (12) yields the assertion. □

4 Typical Cramer's rule ratios

We see in the next section that the proximity can be bounded from above by an expression involving the largest absolute value of the entries of $\boldsymbol{A}_\sigma^{-1}\boldsymbol{A}_{\bar{\sigma}}$, as σ ranges

over all bases of $\boldsymbol{A}$, and $\boldsymbol{A}$ is chosen randomly. Hence, we would like to show that that the largest absolute value of any entry of the matrix $\boldsymbol{A}_\sigma^{-1}\boldsymbol{A}_{\bar{\sigma}}$ is typically not too large, where for our purposes 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 σ. 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

$$\left(\boldsymbol{A}_\sigma^{-1}\boldsymbol{A}_j\right)_i = \frac{\det\left(\boldsymbol{A}_{\sigma - i + j}\right)}{\det\left(\boldsymbol{A}_\sigma\right)}. \tag{13}$$

As before, we let $\nu : \mathscr{G} \to [0, 1]$ denote the $O(n)$-invariant probability measure on $\mathrm{Gr}(d, n)$. The precise statement we show is the following: Fix $\sigma \subseteq [n]$, $i \in \sigma$, $j \in [n] \setminus \sigma$. Then, as a function of a parameter $s > 1$, we have

$$\nu\left(\ker(\boldsymbol{A}) : \boldsymbol{A}_\sigma \text{ is nonsingular}, \left|\left(\boldsymbol{A}_\sigma^{-1}\boldsymbol{A}_j\right)_i\right| > s\right) = \frac{2}{\pi s} + \mathcal{O}\left(s^{-3}\right). \tag{14}$$

The proof proceeds in the three subsections below. First, we get a handle on ν by relating it to another probability distribution, namely the Gaussian distribution γ on the matrix space $\mathbb{R}^{m\times n}$, where the entries are i.i.d. normally distributed with mean 0 and variance 1. This is done via the kernel map, which is introduced in Sect. 4.1 and related to γ in Sect. 4.2. Equation (14) is then derived in Sect. 4.3.

4.1 The real Grassmannian

For a general introduction to matrix groups and Grassmannians, we refer the reader to [4]. 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

$$(\ker(\boldsymbol{A})) \cdot \boldsymbol{U} = \ker(\boldsymbol{A}\boldsymbol{U}). \tag{15}$$

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

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

Let $\mathrm{St}^{m\times n} := \left\{\boldsymbol{A} \in \mathbb{R}^{m\times n} : \mathrm{rank}(\boldsymbol{A}) = m\right\}$. 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:

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

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

$$\boldsymbol{A}\boldsymbol{U}\left(\boldsymbol{A}\boldsymbol{U}\right)^\top = \boldsymbol{A}\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{A}^\top = \boldsymbol{A}\boldsymbol{A}^\top,$$

thus $\boldsymbol{A}$ and $\boldsymbol{AU}$ have the same Gram matrix $\boldsymbol{AA}^\top$, and an $m \times n$ matrix has full-row-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 (15) 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 $\left(\mathbb{R}^{m\times n}, \mathscr{B}(\mathbb{R}^{m\times n}), \gamma\right)$ where $\mathscr{B}(\mathbb{R}^{m\times n})$ is the Borel σ-algebra, and the measure γ is defined so that each $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ has iid $N(0,1)$ entries. In other words, γ 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 $\left(\mathrm{St}^{m\times n}, \mathscr{B}(\mathrm{St}^{m\times n}), \gamma\right)$. 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 γ. 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 if and only if $\ker^{-1}(E)$ is open in $\mathrm{St}^{m\times n}$. Let $\mathscr{G}$ denote the associated Borel σ-algebra. The measure ν on $\mathrm{Gr}\,(d,n)$ is characterized as follows:

Proposition 3 ([13, Corollary 3.1.3]) *The measure* ν *is the unique measure on* $\mathrm{Gr}\,(d,n)$ *satisfying*

$$\begin{aligned} \nu(E \cdot \boldsymbol{U}) &= \nu(E) \text{ for all } E \in \mathscr{G} \text{ and } \boldsymbol{U} \in O(n) \\ \nu(\mathrm{Gr}\,(d,n)) &= 1. \end{aligned} \tag{16}$$

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

Proof We establish the conditions of (16). By surjectivity, and the fact that γ 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

$$\ker^{-1}(E \cdot \boldsymbol{U}) = \ker^{-1}(E) \cdot \boldsymbol{U}. \tag{17}$$

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

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

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

$$\varphi_{\boldsymbol{U}}\,(\boldsymbol{A}) = \boldsymbol{A}\boldsymbol{U} \tag{18}$$

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

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

Now the probability measure γ 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 $\mathcal{K} \in \mathscr{B}\left(\mathbb{R}^{m\times n}\right)$ and $\varphi \in O\,(m\times n)$, we have

$$\gamma\,(\varphi\,(\mathcal{K})) = \gamma\,(\mathcal{K})\,. \tag{19}$$

The same is therefore true for the restricted probability measure γ on $\text{St}^{m\times n}$. It follows that if $\boldsymbol{U} \in O(n)$ and $E \in \mathscr{G}$, then, using (17), (18), and (19), 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

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

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

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

Proposition 5 *The set $\xi_{\sigma,i,j}\,(s)$ is Jordan measurable.*

Proof Let $\xi = \xi_{\sigma,i,j}\,(s)$. We first argue that it suffices to show $\nu\,(\partial\xi) = 0$, where $\partial\xi$ denotes the boundary of ξ. There is a metric on $\text{Gr}\,(d,n)$, which we denote by

δ, whose open balls form a basis for our topology of $\mathrm{Gr}\,(d,n)$. These open balls are defined, for each $\varepsilon > 0$ and d-dimensional subspace V of $\mathbb{R}^n$, as

$$B\,(V,\varepsilon) := \{W \in \mathrm{Gr}\,(d,n) : \delta\,(V,W) < \varepsilon\}\,.$$

Let

$$(\partial\xi)_\varepsilon := \bigcup_{V\in\partial\xi} B\,(V,\varepsilon)\,.$$

Note that $\partial\xi = \cap_{k\geq 1}\,(\partial\xi)_{1/k}$. We have, by the monotone convergence theorem,

$$\lim_{N\to\infty} \nu\left(\bigcap_{k=1}^{N} (\partial\xi)_{1/k}\right) = \nu\left(\bigcap_{k\geq 1} (\partial\xi)_{1/k}\right) = \nu\,(\partial\xi) = 0.$$

In particular, if we now fix some $\varepsilon > 0$, there is some $k \geq 1$ such that $\nu\big((\partial\xi)_{1/k}\big) < \varepsilon$. Observe $\overline{\xi} \cap (\partial\xi)^c_{1/k}$ and $\overline{\xi^c}$ are two disjoint closed sets, where $\overline{X}$, X^c denotes the closure and complement of X in $\mathrm{Gr}\,(d,n)$, respectively. As $\mathrm{Gr}\,(d,n)$ is a metric space it is therefore a normal space, and we may therefore apply Urysohn's lemma [10, Lemma 4.15] to get a continuous function $f_1 : \mathrm{Gr}\,(d,n) \to [0,1]$ such that

$$f_1|_{\overline{\xi}\cap(\partial\xi)^c_{1/k}} = 1 \quad \text{and} \quad f_1|_{\overline{\xi^c}} = 0.$$

Again applying Urysohn's lemma, we also get a function $f_2 : \mathrm{Gr}\,(d,n) \to [0,1]$ such that

$$f_2|_{\overline{\xi^c}\cap(\partial\xi)^c_{1/k}} = 0 \quad \text{and} \quad f_2|_{\overline{\xi}} = 1.$$

Note that by construction, $f_1 \leq \mathbf{1}_\xi \leq f_2$. Furthermore,

$$\int (f_2 - f_1)\,d\nu \leq \nu\big((\partial\xi)_{1/k}\big) < \varepsilon,$$

which establishes the condition of Definition 1.

To conclude the proof, it remains to show $\nu\,(\partial\xi) = 0$. One way to see this is that $\ker^{-1}\left(\overline{\partial\xi}\right)$ is the solution set in $\mathrm{St}^{m\times n}$ to

$$\left(\det \boldsymbol{X}_{\sigma-i+j}\right)^2 - (s\cdot \det \boldsymbol{X}_\sigma)^2 = 0,$$

where $\boldsymbol{X}$ denotes an $m \times n$ matrix of variables. This is an algebraic hypersurface, hence by Proposition 4 we conclude

$$\nu\big(\partial\xi\big) = \gamma\left(\ker^{-1}\left(\overline{\partial\xi}\right)\right) = 0.$$

□

Proposition 6 *For $s > 1$ and σ, i, j as above, we have*

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

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 $\boldsymbol{A}_{\sigma\setminus\{i\}}$, and let ℓ denote the line perpendicular to H. Let $\boldsymbol{u}_\ell$ denote the unit normal vector to H whose first nonzero coordinate is positive. Thus,

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

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 $k \in \sigma\setminus\{i\}$ 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 α.

Now let k be any element of $[n]$ outside of $\sigma\setminus\{i\}$. Since $\boldsymbol{u}_\ell$ depends only on $\boldsymbol{A}_{\sigma\setminus\{i\}}$, 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 σ-algebra generated by ℓ) 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

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

See [9, 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

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

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

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

□

5 Proof of the 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 ω_d denotes the volume of the d-dimensional Euclidean ball of radius 1. 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 Λ is (σ, s, u)-controlled if σ is a basis of $\boldsymbol{A}$ and:

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

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

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

Lemma 3 *If σ is a basis of $\boldsymbol{A}$ and Λ is (σ, s, u)-controlled, then for all $\boldsymbol{x}^* \in \mathcal{S}_\sigma(\Lambda)$ we have*

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

Proof Let $\boldsymbol{b} = \boldsymbol{A}\boldsymbol{x}^*$, let $B = B_2^n \cap \Lambda_{\mathbb{R}}$, and let μ denote the covering radius of B with respect to Λ. 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 $\boldsymbol{x}^* \in \mathcal{S}_\sigma(\Lambda)$ 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 $\mu B + \boldsymbol{v} \subseteq \mathcal{P}(\boldsymbol{A}, \boldsymbol{b})$.

Let $\mathbf{g} \in (\Lambda_{\mathbb{R}} + \boldsymbol{x}^*) \cap \mathbb{Z}^n$. Since μ is the covering radius of B with respect to Λ, there exists $\boldsymbol{z}^* \in (\Lambda + \mathbf{g}) \cap (\mu B + \boldsymbol{v})$ such that

$$\left\|\boldsymbol{x}^* - \boldsymbol{z}^*\right\|_2 \leq \left\|\boldsymbol{x}^* - \boldsymbol{v}\right\|_2 + \left\|\boldsymbol{v} - \boldsymbol{z}^*\right\|_2 \leq \mu \left\|\tilde{\boldsymbol{w}}\right\|_2 + \mu. \tag{20}$$

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

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

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

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

□

Proof Let Λ 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 3. We have

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

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

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

By definition, this probability is at most

$$\mathbf{P}\left(\max_{i\in[d-1]}\left\{\frac{\lambda_{i+1}(\Lambda)}{\lambda_i(\Lambda)}\right\} \geq (\tilde{\omega}_d u)^{2/(d-1)}\right) + \sum_{\substack{i\in\sigma \\ j\notin\sigma}} \mathbf{P}\left(\sigma \text{ basis of } \boldsymbol{A}, \left(\boldsymbol{A}_\sigma^{-1}\boldsymbol{A}_j\right)_i \geq s\right). \tag{21}$$

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 2 and Proposition 6, for T sufficiently large, we may estimate up to constants the quantity (21) by

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

□

Acknowledgements Marcel Celaya 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). The authors wish to thank the anonymous referees for their helpful comments and suggestions.

Funding Open access funding provided by Swiss Federal Institute of Technology Zurich

References

1. Aliev, I., Henk, M.: Feasibility of integer knapsacks. SIAM J. Optim. **20**(6), 2978–2993 (2010)
2. Aliev, I., Henk, M., Oertel, T.: Distances to lattice points in knapsack polyhedra. Math. Program. **182**((1–2, Ser. A)), 175–198 (2020)
3. Artstein-Avidan, S., Giannopoulos, A., Milman, V.D.: Asymptotic Geometric Analysis: Part I. Mathematical Surveys and Monographs, vol. 202. American Mathematical Society, Providence (2015)
4. Baker, A.: Matrix Groups: An Introduction to Lie Group Theory. Springer Undergraduate Mathematics Series, Springer, London (2003)
5. Borst, S., Dadush, D., Huiberts, S., Tiwari, S.: On the integrality gap of binary integer programs with gaussian data. In: Integer Programming and Combinatorial Optimization. Lecture Notes in Computer Science, vol. 12707, pp. 427–442. Springer, Cham (2021)
6. Celaya, M., Henk, M.: Proximity bounds for random integer programs. In: Integer Programming and Combinatorial Optimization. Lecture Notes in Computer Science, vol. 12707, pp. 413–426. Springer, Cham (2021)
7. Cook, W., Gerards, A.M.H., Schrijver, A., Tardos, É.: Sensitivity theorems in integer linear programming. Math. Program. **34**(3), 251–264 (1986)
8. Eisenbrand, F., Weismantel, R.: Proximity results and faster algorithms for integer programming using the Steinitz lemma. ACM Trans. Algorithms **16**(1), 1–14 (2019)
9. Feller, V., Feller, W.: An Introduction to Probability Theory and Its Applications. A Wiley Publication in Mathematical Statistics, vol. 1. Wiley, Hoboken (1968)
10. Folland, G.B.: Real Analysis: Modern Techniques and Their Applications. Pure and Applied Mathematics: A Wiley Series of Texts, Monographs and Tracts, Wiley, Hoboken (1999)
11. Gomory, R.E.: On the relation between integer and noninteger solutions to linear programs. Proc. Natl. Acad. Sci. U.S.A. **53**(2), 260 (1965)
12. Gruber, P.M.: Convex and Discrete Geometry. Fundamental Principles of Mathematical Sciences, vol. 336. Springer, Berlin (2007)
13. Krantz, S.G., Parks, H.R.: Geometric Integration Theory. Cornerstones. Birkhäuser, Boston (2008)
14. Oertel, T., Paat, J., Weismantel, R.: The distributions of functions related to parametric integer optimization. SIAM J. Appl. Algebra Geom. **4**(3), 422–440 (2020)
15. Schmidt, W.M.: The distribution of sublattices of $\mathbf{Z}^m$. Monatshefte Math. **125**(1), 37–81 (1998)