Integer Programming Notes / Winter 21/22

MARTIN KOUTECKÝ, Charles University, Czech Republic

1 INTEGER PROGRAMMING IN GENERAL

Our focus is on the integer (linear) programming problem in standard form

$$\min \{f(\mathbf{x}) \mid A\mathbf{x} = \mathbf{b},\ \mathbf{l} \le \mathbf{x} \le \mathbf{u},\ \mathbf{x} \in \mathbb{Z}^n\},\ \text{and} \tag{IP}$$

$$\min \{\mathbf{w}\mathbf{x} \mid A\mathbf{x} = \mathbf{b},\ \mathbf{l} \le \mathbf{x} \le \mathbf{u},\ \mathbf{x} \in \mathbb{Z}^n\}, \tag{ILP}$$

with A an integer $m \times n$ matrix, $f : \mathbb{R}^n \to \mathbb{R}$ a separable convex function, $\mathbf{b} \in \mathbb{Z}^m$, and $\mathbf{l}, \mathbf{u} \in (\mathbb{Z} \cup \{\pm\infty\})^n$. (IP) is well-known to be strongly NP-hard already in the special case (ILP) when $f(\mathbf{x}) = \mathbf{w}\mathbf{x}$ is a linear objective function for some vector $\mathbf{w} \in \mathbb{Z}^n$. In this course, we will cover some important, broad, natural. and useful conditions under which (IP) can be solved in polynomial time.

Notation

We write vectors in boldface (e.g., $\mathbf{x}, \mathbf{y}$) and their entries in normal font (e.g., the i-th entry of $\mathbf{x}$ is x_i). For positive integers $m \le n$ we set $[m, n] := \{m, \ldots, n\}$ and $[n] := [1, n]$, and we extend this notation for vectors: for $\mathbf{l}, \mathbf{u} \in \mathbb{Z}^n$ with $\mathbf{l} \le \mathbf{u}$, $[\mathbf{l}, \mathbf{u}] := \{\mathbf{x} \in \mathbb{Z}^n \mid \mathbf{l} \le \mathbf{x} \le \mathbf{u}\}$. If A is a matrix, $A_{i,j}$ denotes the j-th coordinate of the i-th row, $A_{i,\bullet}$ denotes the i-th row and $A_{\bullet,j}$ denotes the j-th column. We use $\log := \log_2$. For an integer $a \in \mathbb{Z}$, we denote by $\langle a \rangle := 1 + \lceil \log(|a| + 1) \rceil$ the binary encoding length of a; we extend this notation to vectors, matrices and tuples of these objects. For example, $\langle A, \mathbf{b} \rangle = \langle A \rangle + \langle \mathbf{b} \rangle$, and $\langle A \rangle = \sum_{i,j} \langle A_{i,j} \rangle$.

2 FIXED DIMENSION

(IP) can be solved in time $g(n)poly(n, L)$ for some function g, and this goes back to the work of Lenstra [8]. The best current bound is $g(n) = O(n)^n$ and is due to Dadush [1]. The algorithm even applies to the case where f is general convex (non-separable), and where $\mathbf{x}$ belongs to some convex body $K \subseteq \mathbb{R}^n$.

We will sketch the main ideas of the proof, but don't pretend to give all details. Define the *width of K along a direction* $\mathbf{d} \in \mathbb{Z}^n$ to be

$$w_{\mathbf{d}}(K) = \max\{\mathbf{d}\mathbf{x} \mid \mathbf{x} \in K\} - \min\{\mathbf{d}\mathbf{x} \mid \mathbf{x} \in K\} \ .$$

If the max or min does not exist, we define the width to be infinity. The *width of K* is defined as the smallest width over all non-zero directions; notice that we are taking the directions over all integer vectors in order to avoid silly issues like being able to get very small width by taking very small (non-integral) $\mathbf{d}$:

$$w(K) = \min_{\mathbf{d} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}} w_{\mathbf{d}}(K) \ .$$

A $\mathbf{d}$ which attains the minimum above is called a *flat direction of K*. The algorithm relies on the following deep and famous result:

Proposition 1 (Khinchine's Flatness Theorem). *Let $K \subseteq \mathbb{R}^n$ be a convex body. Then either K contains a lattice point (i.e., $K \cap \mathbb{Z}^n \neq \emptyset$), or $w(K) \le \omega(n)$ where $\omega(n)$ is some constant only depending on n.*

Author's address: Martin Koutecký, Charles University, Prague, Czech Republic, koutecky@iuuk.mff.cuni.cz.

(It could be that K is flat *and* contains an integer point, and this doesn't bother us.)

We focus on solving feasibility, that is, deciding $K \cap \mathbb{Z}^n \neq \emptyset$; optimization can be handled by doing a binary search over the objective and then adding this objective bound into the set of constraints. Specifically, if our guess on the objective is T, we want to enforce a constraint $f(\mathbf{x}) \leq T$, and because f is convex, this is a convex constraint and thus $K' = K \cap \{\mathbf{x} \mid f(\mathbf{x}) \leq T\}$ is a convex set and we solve feasibility for K' instead of optimization over K.

The main idea of the algorithm is this. We compute a flat direction $\mathbf{d}$ of K (which is not an easy problem but is known to be doable so we treat it here as an oracle call). If we see that $w(K) > \omega(n)$, we know that K contains an integer point and we are done. Otherwise, $w(K) \leq \omega(n)$. This means we can branch into at most $\omega(n)$ lower-dimensional slices of K and solve the problem inductively in each of them. Because the dimension drops by at least one in each branching, the branching tree has at most n levels, and because we branch into at most $\omega(n)$ slices, the degree of the tree is at most $\omega(n)$, so altogether the tree has at most $\omega(n)^n$ nodes.

What does this branching look like in detail? If K contains an integer point, then it must lie on one of the hyperplanes

$$\mathbf{dx} = \delta, \text{ where } \delta \in [\min\{\mathbf{dx} \mid \mathbf{x} \in K\}, \max\{\mathbf{dx} \mid \mathbf{x} \in K\}] \ .$$

The rest of the work is that we need to transform the set $K \cap \{\mathbf{x} \mid \mathbf{dx} = \delta\}$ which is less than n-dimensional but lives in n dimensions into a set $K' \subseteq \mathbb{R}^{n-1}$ which is integer feasible iff K is, and then call the algorithm on K'.

3 VARIABLE DIMENSION

Some more preliminaries are in order now.

For a function $f : \mathbb{Z}^n \to \mathbb{Z}$ and two vectors $\mathbf{l}, \mathbf{u} \in \mathbb{Z}^n$, we define $f_{\max}^{[\mathbf{l},\mathbf{u}]} := \max_{\mathbf{x},\mathbf{x}' \in [\mathbf{l},\mathbf{u}]} |f(\mathbf{x}) - f(\mathbf{x}')|$; if $[\mathbf{l}, \mathbf{u}]$ is clear from the context we omit it and write just $f_{\max}$. We assume that $f : \mathbb{R}^n \to \mathbb{R}$ is a separable convex function, i.e., it can be written as $f(\mathbf{x}) = \sum_{i=1}^{n} f_i(x_i)$ where f_i is a convex function of one variable, for each $i \in [n]$. Moreover, we require that for each $\mathbf{x} \in \mathbb{Z}^n$, $f(\mathbf{x}) \in \mathbb{Z}$. We assume f is given by a comparison oracle. We use ω to denote the smallest number such that matrix multiplication of $n \times n$ matrices can be performed in time $O(n^\omega)$. We say that a system of equations $A\mathbf{x} = \mathbf{b}$ is *pure* if the rows of A are linearly independent. The next statement follows easily by Gaussian elimination, hence we assume $m \leq n$ throughout the paper.

Proposition 2 (Purification [5, Theorem 1.4.8]). *Given $A \in \mathbb{Z}^{m \times n}$ and $\mathbf{b} \in \mathbb{Z}^m$ one can in time $O(\min\{n, m\}nm)$ either declare $A\mathbf{x} = \mathbf{b}$ infeasible, or output a pure equivalent subsystem $A'\mathbf{x} = \mathbf{b}'$.*

The goal of this section is to prove the following theorem:

Theorem **??**. *There is a computable function g such that (IP) can be solved in time*

$$g(\|A\|_\infty, \min\{\mathrm{td}_P(A), \mathrm{td}_D(A)\}) \cdot n^2 \log \|\mathbf{u} - \mathbf{l}, \mathbf{b}\|_\infty \log(2f_{\max}) + O(n^\omega \langle A \rangle)$$

In Sections 3.1-3.2 we shall develop the necessary ingredients to prove this theorem. Then, we will conclude in Section **??** by providing its proof which puts these ingredients together.

3.1 Introduction to Iterative Augmentation

Let us introduce Graver bases and discuss how they are used for optimization. We define a partial order $\sqsubseteq$ on $\mathbb{R}^n$ as follows: for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ we write $\mathbf{x} \sqsubseteq \mathbf{y}$ and say that $\mathbf{x}$ is *conformal* to $\mathbf{y}$ if, for each $i \in [n]$, $x_i y_i \geq 0$ (that is, $\mathbf{x}$ and $\mathbf{y}$ lie in the same orthant) and $|x_i| \leq |y_i|$. For a matrix $A \in \mathbb{Z}^{m \times n}$ we write $\ker_{\mathbb{Z}}(A) = \{\mathbf{x} \in \mathbb{Z}^n \mid A\mathbf{x} = \mathbf{0}\}$. It is well known that every subset of $\mathbb{Z}^n$ has finitely many $\sqsubseteq$-minimal elements [3].

Definition 3 (Graver basis [4]). The *Graver basis* of an integer $m \times n$ matrix A is the finite set $\mathcal{G}(A) \subset \mathbb{Z}^n$ of $\sqsubseteq$-minimal elements in $\ker_{\mathbb{Z}}(A) \setminus \{\mathbf{0}\}$.

One important property of $\mathcal{G}(A)$ is as follows:

Lemma 4 (Positive Sum Property [9, Lemma 3.4]). *Let $A \in \mathbb{Z}^{m\times n}$. For any $\mathbf{x} \in \ker_{\mathbb{Z}}(A)$, there exists an $n' \leq 2n-1$ and a decomposition $\mathbf{x} = \sum_{j=1}^{n'} \lambda_j \mathbf{g}_j$ with $\lambda_j \in \mathbb{N}$ and $\mathbf{g}_j \in \mathcal{G}(A)$ for each $j \in [n']$, and with $\mathbf{g}_j \sqsubseteq \mathbf{x}$, i.e., all $\mathbf{g}_j$ belonging to the same orthant as $\mathbf{x}$.*

Proof. Let G be a matrix whose columns are $\mathbf{g} \in \mathcal{G}(A)$ such that $\mathbf{g} \sqsubseteq \mathbf{x}$. Consider the following LP in variables $\mathbf{y} \in \mathbb{R}^{|G|}$:

$$\begin{aligned} \max \sum_{\mathbf{g}} y_{\mathbf{g}} \\ G\mathbf{y} = \mathbf{x} \\ \mathbf{y} \geq \mathbf{0} \end{aligned}$$

There is a basic optimal solution $\mathbf{y}^*$ and from LP theory we know that, because there are n equality constraints and only non-negativity constraints besides that, $|\operatorname{supp}(\mathbf{y}^*)| \leq n$. We will define the coefficient vector $\boldsymbol{\lambda}$ in two phases. In the first phase, let $\boldsymbol{\lambda} = \lfloor \mathbf{y}^* \rfloor$. Recall that $\{\mathbf{y}^*\}$ is the fractional part of $\mathbf{y}^*$. Observe that $G\{\mathbf{y}^*\}$ is an integer vector, because it is $G\mathbf{y}^* - G\lfloor \mathbf{y}^* \rfloor$ which is a difference of two integer vectors. Thus, $\{\mathbf{y}^*\}$ describes a decomposition of $\bar{\mathbf{x}} := \mathbf{x} - G\lfloor \mathbf{y}^* \rfloor \in \operatorname{Ker}_{\mathbb{Z}}(A)$. Moreover, $\{\mathbf{y}^*\}$ is a decomposition maximizing the ℓ_1-norm, which is the objective of the above LP. (The fact that $\{\mathbf{y}^*\}$ is a fractional decomposition of $\bar{\mathbf{x}}$ maximizing ℓ_1-norm is easy to see by contradiction: if there was a better decomposition $\mathbf{y}'$ of $\bar{\mathbf{x}}$, one could use it to get a better decomposition $\mathbf{y}' + \lfloor \mathbf{y}^* \rfloor$ of $\mathbf{x}$, but $\mathbf{y}^*$ was assumed to be maximum.) Finally, we have that $\|\{\mathbf{y}^*\}\|_1 < n$ because it is a sum of at most n numbers, each strictly smaller than 1. Now consider an optimal *integer* decomposition of $\bar{\mathbf{x}}$, i.e., a non-negative vector $\bar{\mathbf{y}} \in \mathbb{Z}^n$ satisfying $G\bar{\mathbf{y}} = \bar{\mathbf{x}}$. It cannot have a larger ℓ_1-norm than $\{\mathbf{y}^*\}$ because $\{\mathbf{y}^*\}$ is an optimum of the continuous relaxation, thus $\|\bar{\mathbf{y}}\|_1 \leq \|\{\mathbf{y}^*\}\|_1 < n$, that is, $\|\bar{\mathbf{y}}\|_1 \leq n-1$. This is the second phase: update $\boldsymbol{\lambda}_: = \boldsymbol{\lambda} + \bar{\mathbf{y}}$. We have $|\operatorname{supp}(\boldsymbol{\lambda})| \leq |\operatorname{supp}(\mathbf{y}^*)| + |\operatorname{supp}(\bar{\mathbf{y}})| \leq n + (n-1) = 2n-1$. □

In fact, the Lemma holds with a better constant $2n-2$, and we will use this bound in the sequel, although this has no asymptotic significance for us.

Proposition 5 (Positive Sum Property [9, Lemma 3.4]). *For any $\mathbf{x} \in \ker_{\mathbb{Z}}(A)$, there exists an $n' \leq 2n-2$ and a decomposition $\mathbf{x} = \sum_{j=1}^{n'} \lambda_j \mathbf{g}_j$ with $\lambda_j \in \mathbb{N}$ and $\mathbf{g}_j \in \mathcal{G}(A)$ for each $j \in [n']$, and with $\mathbf{g}_j \sqsubseteq \mathbf{x}$, i.e., all $\mathbf{g}_j$ belonging to the same orthant as $\mathbf{x}$.*

We say that $\mathbf{x} \in \mathbb{Z}^n$ is *feasible* for (IP) if $A\mathbf{x} = \mathbf{b}$ and $\mathbf{l} \leq \mathbf{x} \leq \mathbf{u}$. Let $\mathbf{x}$ be a feasible solution for (IP). We call $\mathbf{g}$ a *feasible step* if $\mathbf{x} + \mathbf{g}$ is feasible for (IP). Further, call a feasible step $\mathbf{g}$ *augmenting* if $f(\mathbf{x}+\mathbf{g}) < f(\mathbf{x})$. An important implication of Proposition 5 is that if *any* augmenting step exists, then there exists one in $\mathcal{G}(A)$ [2, Lemma 3.3.2].

An augmenting step $\mathbf{g}$ and a *step length* $\lambda \in \mathbb{N}$ form an $\mathbf{x}$-*feasible step pair* with respect to $\mathbf{x}$ if $\mathbf{l} \leq \mathbf{x} + \lambda\mathbf{g} \leq \mathbf{u}$. An augmenting step $\mathbf{h}$ is a *Graver-best step* for $\mathbf{x}$ if $f(\mathbf{x}+\mathbf{h}) \leq f(\mathbf{x}+\lambda\mathbf{g})$ for all $\mathbf{x}$-feasible step pairs $(\mathbf{g}, \lambda) \in \mathcal{G}(A) \times \mathbb{N}$. A slight relaxation of a Graver-best step is a halfling: an augmenting step $\mathbf{h}$ is a *halfling* for $\mathbf{x}$ if $f(\mathbf{x}) - f(\mathbf{x}+\mathbf{h}) \geq \frac{1}{2}(f(\mathbf{x}) - f(\mathbf{x}+\lambda\mathbf{g}))$ for all $\mathbf{x}$-feasible step pairs $(\mathbf{g}, \lambda) \in \mathcal{G}(A) \times \mathbb{N}$. A *halfling augmentation procedure* for (IP) with a given feasible solution $\mathbf{x}_0$ works as follows. Let $i := 0$.

(1) If there is no halfling for $\mathbf{x}_i$, return it as optimal.
(2) If a halfling $\mathbf{h}_i$ for $\mathbf{x}_i$ exists, set $\mathbf{x}_{i+1} := \mathbf{x}_i + \mathbf{h}_i$, $i := i+1$, and go to 1.

We assume that the bounds $\mathbf{l}, \mathbf{u}$ are finite. Since there are several approaches how to achieve this, we postpone the discussion on dealing with infinite bounds to Section ??.

Lemma 6 (Halfling convergence). *Given a feasible solution $\mathbf{x}_0$ for (IP), the halfling augmentation procedure finds an optimum of (IP) in at most $3n \log\left(f(\mathbf{x}_0) - f(\mathbf{x}^*)\right) \leq 3n \log\left(f_{\max}^{[\mathbf{l},\mathbf{u}]}\right)$ steps.*

Before we prove the lemma we need a useful proposition about separable convex functions:

Proposition 7 (Separable convex superadditivity [2, Lemma 3.3.1]). *Let $f(\mathbf{x}) = \sum_{i=1}^n f_i(x_i)$ be separable convex, let $\mathbf{x} \in \mathbb{R}^n$, and let $\mathbf{g}_1, \ldots, \mathbf{g}_k \in \mathbb{R}^n$ be vectors that are pairwise conformal. Then*

$$f\left(\mathbf{x} + \sum_{j=1}^{k} \alpha_j \mathbf{g}_j\right) - f(\mathbf{x}) \geq \sum_{j=1}^{k} \alpha_j \left(f(\mathbf{x} + \mathbf{g}_j) - f(\mathbf{x})\right) \tag{1}$$

for arbitrary integers $\alpha_1, \ldots, \alpha_k \in \mathbb{N}$.

The essence of this proposition is that if multiple steps take us from $\mathbf{x}_0$ to $\mathbf{x}^*$, then the sum of their improvements considered individually with respect to $\mathbf{x}_0$ is *at least* the improvement when we take them together, i.e., the difference $f(\mathbf{x}) - f(\mathbf{x}^*)$. (This is tricky to read correctly because "improvement" is a negative term, because we are minimizing.) An illustrative example is $f(x) = x^2$: if we are at a point $x = 2$, then moving by 1 closer to the origin improves the objective by 3 (decrease from 4 to 1), but moving by another 1 only improves it by 1 (decrease from 1 to 0). So if we think of the path from 2 to 0 as two steps by 1, when we consider the total of the progress each step would achieve individually with respect to the initial point 2, we get $3 + 3 = 6$, but taken together, the steps only achieve the progress of 4. Essentially, the contribution of each step considered in the sequence is at most the contribution of each step considered individually with respect to the initial point.

Proof of Lemma 6. Let $\mathbf{x}^*$ be an optimal solution of (IP). By Proposition 5 we may write $\mathbf{x}^* - \mathbf{x}_0 = \sum_{j=1}^{n'} \lambda_j \mathbf{g}_j$ such that $\mathbf{g}_j \sqsubseteq \mathbf{x}^* - \mathbf{x}_0$ for all $j \in [n']$, and $n' \leq 2n - 2$. We apply Proposition 7 to $\mathbf{x}_0$ and the n' vectors $\lambda_j \mathbf{g}_j$ with $\alpha_j := 1$, so by (1) we have

$$0 \geq f(\mathbf{x}^*) - f(\mathbf{x}_0) = f\left(\mathbf{x}_0 + \sum_{j=1}^{n'} \lambda_j \mathbf{g}_j\right) - f(\mathbf{x}_0) \geq \sum_{j=1}^{n'} \left(f(\mathbf{x}_0 + \lambda_j \mathbf{g}_j) - f(\mathbf{x}_0)\right),$$

and multiplying by -1 gives $0 \leq f(\mathbf{x}_0) - f(\mathbf{x}^*) \leq \sum_{j=1}^{n'} (f(\mathbf{x}_0) - f(\mathbf{x}_0 + \lambda_j \mathbf{g}_j))$. By an averaging argument, there must exist an index $\ell \in [n']$ such that

$$f(\mathbf{x}_0) - f(\mathbf{x}_0 + \lambda_\ell \mathbf{g}_\ell) \geq \frac{1}{n'}\left(f(\mathbf{x}_0) - f(\mathbf{x}^*)\right) \geq \frac{1}{n'} f_{\max} \quad . \tag{2}$$

Consider a halfling $\mathbf{h}$ for $\mathbf{x}_0$: by definition, it satisfies $f(\mathbf{x}_0) - f(\mathbf{x}_0 + \mathbf{h}) \geq \frac{1}{2}(f(\mathbf{x}_0) - f(\mathbf{x}_0 + \lambda_i \mathbf{g}_i))$. Say that the halfling augmentation procedure required s iterations. For $i \in [s-1]$ we have that

$$f(\mathbf{x}_i) - f(\mathbf{x}^*) \leq \left(1 - \frac{1}{4n-4}\right)\left(f(\mathbf{x}_{i-1}) - f(\mathbf{x}^*)\right) = \frac{4n-5}{4n-4}\left(f(\mathbf{x}_{i-1}) - f(\mathbf{x}^*)\right)$$

and, by repeated application of the above,

$$f(\mathbf{x}_i) - f(\mathbf{x}^*) \leq \left(\frac{4n-5}{4n-4}\right)^i \left(f(\mathbf{x}_0) - f(\mathbf{x}^*)\right) \quad .$$

Since i is not the last iteration, $f(\mathbf{x}_i) - f(\mathbf{x}^*) \geq 1$ by the integrality of f. Take $t := 4n - 4$ and compute

$$
\begin{aligned}
1 &\leq \left(\frac{t-1}{t}\right)^i (f(\mathbf{x}_0) - f(\mathbf{x}^*)) && / \ln(\cdot) \\
0 &\leq i \ln\left(\frac{t-1}{t}\right) + \ln(f(\mathbf{x}_0) - f(\mathbf{x}^*)) && / -i\ln\left(\frac{t-1}{t}\right) \\
-i\ln\left(\frac{t-1}{t}\right) = i\ln\left(\frac{t}{t-1}\right) &\leq \ln(f(\mathbf{x}_0) - f(\mathbf{x}^*)) && / : \ln\left(\frac{t-1}{t}\right) \\
i &\leq \left(\ln\left(\frac{t}{t-1}\right)\right)^{-1} \ln(f(\mathbf{x}_0) - f(\mathbf{x}^*))
\end{aligned}
$$

Now Taylor expansion gives for $t \geq 3$

$$
\ln\left(1 + \frac{1}{t-1}\right) \geq \frac{1}{t-1} - \frac{1}{2(t-1)^2} = \frac{2t-3}{2t^2 - 4t + 2},
$$

so

$$
\left(\ln\left(1 + \frac{1}{t-1}\right)\right)^{-1} \leq \frac{2t^2 - 4t + 2}{2t - 3},
$$

which is bounded above by t for all $t \geq 2$ since $t(2t-3) = 2t^2 - 3t \geq 2t^2 - 4t + 2$ for all $t \geq 2$. Plugging back $t := 4n - 4$ we get that for all $n \geq 2$ we have $t \geq 3$ and hence

$$
i \leq (4n-4)\ln(f(\mathbf{x}_0) - f(\mathbf{x}^*)) = (4n-4) \cdot \ln 2 \cdot \log_2(f(\mathbf{x}_0) - f(\mathbf{x}^*)),
$$

and the number of iterations is at most one unit larger. Since $f(\mathbf{x}_0) - f(\mathbf{x}^*) \leq f_{\max}$ and $\ln(2) = 0.693147\cdots \leq 3/4$, we have that the number of iterations is at most $3n\log(f_{\max})$. □

Clearly it is now desirable to show how to find halflings quickly. The following lemma will be helpful in that regard.

Lemma 8 (Powers of Two). *Let $\Gamma_2 = \{1, 2, 4, 8, \dots\}$ and $\mathbf{x}$ be a feasible solution of* (IP). *If $\mathbf{h}$ satisfies $f(\mathbf{x} + \mathbf{h}) \leq f(\mathbf{x} + \lambda\mathbf{g})$ for each $\mathbf{x}$-feasible step pair $(\mathbf{g}, \lambda) \in \mathcal{G}(A) \times \Gamma_2$, then $\mathbf{h}$ is a halfling.*

Proof. Consider any Graver-best step pair $(\mathbf{g}^*, \lambda^*) \in \mathcal{G}(A) \times \mathbb{N}$, let $\lambda := 2^{\lfloor \log \lambda^* \rfloor}$, and choose $\frac{1}{2} < \gamma \leq 1$ in such a way that $\lambda = \gamma\lambda^*$. Convexity of f yields

$$
\begin{aligned}
f(\mathbf{x}_0) - f(\mathbf{x}_0 + \lambda\mathbf{g}^*) &\geq f(\mathbf{x}_0) - [(1-\gamma)f(\mathbf{x}_0) + \gamma f(\mathbf{x}_0 + \lambda^*\mathbf{g}^*)] \\
&= \gamma(f(\mathbf{x}_0) - f(\mathbf{x}_0 + \lambda^*\mathbf{g}^*)) \\
&\geq \frac{1}{2}(f(\mathbf{x}_0) - f(\mathbf{x}_0 + \lambda^*\mathbf{g}^*)) \quad .
\end{aligned}
$$

This shows that $\lambda\mathbf{g}^*$ is a halfling, and by the definition of $\mathbf{h}$, $f(\mathbf{x} + \mathbf{h}) \leq f(\mathbf{x} + \lambda\mathbf{g}^*)$ and thus $\mathbf{h}$ is a halfling as well. □

This makes it clear that the main task is to find, for each $\lambda \in \Gamma_2$, a step $\mathbf{h}$ which is at least as good as any feasible $\lambda\mathbf{g}$ with $\mathbf{g} \in \mathcal{G}(A)$. We need the notion of a *best* solution:

Definition 9 (S-best solution). Let $S, P \subseteq \mathbb{R}^n$. We say that $\mathbf{x}^* \in P$ is a solution of

$$
S\text{-best}\,\{f(\mathbf{x}) \mid \mathbf{x} \in P\} \qquad (S\text{-best})
$$

if $f(\mathbf{x}^*) \leq \min\{f(\mathbf{x}) \mid \mathbf{x} \in P \cap S\}$. If $P \cap S$ is empty, we say S-best $\{f(\mathbf{x}) \mid \mathbf{x} \in P\}$ has no solution.

In other words, $\mathbf{x}^*$ has to belong to P and be at least as good as any point in $P \cap S$. Note that to define the notion of an S-best solution to be a "no solution" if $P \cap S = \emptyset$ might look unnatural as one might require *any* $\mathbf{x} \in P$ if $P \cap S = \emptyset$. However, this would make (S-best) as hard as finding some $\mathbf{x} \in P$ (just take $S = \emptyset$), but intuitively (S-best) should be an easier problem. The following is a central notion.

Definition 10 (Augmentation IP). For an (IP) instance $(A, f, \mathbf{b}, \mathbf{l}, \mathbf{u})$, its feasible solution $\mathbf{x} \in \mathbb{Z}^n$, and an integer $\lambda \in \mathbb{N}$, the *Augmentation IP* problem is to solve

$$\mathcal{G}(A)\text{-best}\{f(\mathbf{x} + \lambda \mathbf{g}) \mid A\mathbf{g} = \mathbf{0},\ \mathbf{l} \le \mathbf{x} + \lambda\mathbf{g} \le \mathbf{u},\ \mathbf{g} \in \mathbb{Z}^n\} \ . \tag{AugIP}$$

Let $(A, f, \mathbf{b}, \mathbf{l}, \mathbf{u})$ be an instance of (IP), $\mathbf{x}$ a feasible solution, and $\lambda \in \mathbb{N}$. We call the pair $(\mathbf{x}, \lambda)$ an (AugIP) *instance* for $(A, f, \mathbf{b}, \mathbf{l}, \mathbf{u})$. If clear from the context, we omit the (IP) instance $(A, f, \mathbf{b}, \mathbf{l}, \mathbf{u})$.

By Lemma 8 we obtain a halfling by solving (AugIP) for each $\lambda \in \Gamma_2$ and picking the best solution. Given an initial feasible solution $\mathbf{x}_0$ and a fast algorithm for (AugIP), we can solve (IP) quickly:

Lemma 11 (((AugIP) and $\mathbf{x}_0$) $\implies$ (IP)). *Given an initial feasible solution $\mathbf{x}_0$ to (IP), (IP) can be solved by solving*

$$3n(\log \|\mathbf{u} - \mathbf{l}\|_\infty + 1) \log(f(\mathbf{x}_0) - f(\mathbf{x}^*)) \le 3n(\log \|\mathbf{u} - \mathbf{l}\|_\infty + 1) \log\left(f_{\max}^{[\mathbf{l},\mathbf{u}]}\right)$$

instances of (AugIP), where $\mathbf{x}^*$ is any optimum of (IP).

Proof. Observe that no $\lambda \in \Gamma_2 = \{1, 2, 4, \dots\}$ greater than $\|\mathbf{u} - \mathbf{l}\|_\infty$ results in a non-zero $\mathbf{x}$-feasible step pair. Thus, by Lemma 8, to compute a halfling for $\mathbf{x}$ it suffices to solve (AugIP) for all $\lambda \in \Gamma_2$, $\lambda \le \|\mathbf{u} - \mathbf{l}\|_\infty$, and there are at most $\log \|\mathbf{u} - \mathbf{l}\|_\infty + 1$ of these. By Lemma 6, $3n \log(f(\mathbf{x}_0) - f(\mathbf{x}^*)) \le 3n \log(f_{\max})$ halfling augmentations suffice and we are done. □

Feasibility. Our goal now is to satisfy the requirement of an initial solution $\mathbf{x}_0$.

Lemma 12 ((AugIP) $\implies$ $\mathbf{x}_0$). *Given an instance of (IP), it is possible to compute a feasible solution $\mathbf{x}_0$ for (IP) or decide that (IP) is infeasible by solving $O(n \log(\|A, \mathbf{b}, \mathbf{l}, \mathbf{u}\|_\infty)^2)$ many (AugIP) instances, plus $O(n^\omega)$ time needed to compute an integral solution of $A\mathbf{z} = \mathbf{b}$. Moreover, $\langle \mathbf{x}_0 \rangle \le \text{poly}\langle \mathbf{b} \rangle$.*

Proof. We first compute an integer solution to the system of equations $A\mathbf{z} = \mathbf{b}$. This can be done by computing the Hermite normal form of A in time $O(n^{\omega-1}m) \le O(n^\omega)$ [12] (using $m \le n$). Then either we conclude that there is no integer solution to $A\mathbf{z} = \mathbf{b}$ and hence (IP) is infeasible, or we find a solution $\mathbf{z} \in \mathbb{Z}^n$ with encoding length polynomially bounded in the encoding length of $A, \mathbf{b}$.

Next, we will solve an auxiliary IP. Define new relaxed bounds by

$$\hat{l}_i := \min\{l_i, z_i\}, \quad \hat{u}_i := \max\{u_i, z_i\}, \quad i \in [n],$$

and define an objective function $\hat{f} := \sum_{i=1}^n \hat{f}_i$ as, for each $i \in [n]$, $\hat{f}_i(x_i) := \text{dist}(x_i, [l_i, u_i])$, which is 0 if $x_i \in [l_i, u_i]$ and $\max\{l_i - x_i, x_i - u_i\}$ otherwise. This function has at most three linear pieces, the first decreasing, the second constantly zero, and the third increasing, and thus each $\hat{f}_i$ is convex and $\hat{f}$ is separable convex. Moreover, a solution $\mathbf{x}$ has $\hat{f}(\mathbf{x}) = 0$ if and only if $\mathbf{l} \le \mathbf{x} \le \mathbf{u}$.

By Lemma 6, an optimum $\mathbf{x}_0$ of $\min\left\{\hat{f}(\mathbf{x}) \,\middle|\, A\mathbf{x} = \mathbf{b},\ \hat{\mathbf{l}} \le \mathbf{x} \le \hat{\mathbf{u}},\ \mathbf{x} \in \mathbb{Z}^n\right\}$ can be computed by solving $3n(\log \|\hat{\mathbf{u}} - \hat{\mathbf{l}}\| + 1) \log\left(\hat{f}_{\max}^{[\hat{\mathbf{l}},\hat{\mathbf{u}}]}\right)$ instances of (AugIP). Since $\|\hat{\mathbf{l}}, \hat{\mathbf{u}}\|_\infty$ is polynomially bounded in $\|A, \mathbf{b}\|_\infty$ and $\|\mathbf{l}, \mathbf{u}\|_\infty$ and, by definition of $\hat{f}$, $\hat{f}_{\max}^{[\hat{\mathbf{l}},\hat{\mathbf{u}}]}$ is bounded by $n \cdot \|\hat{\mathbf{l}}, \hat{\mathbf{u}}\|_\infty$, we have that the number of times we have to solve (AugIP) is bounded by $O(n \log(\|A, \mathbf{b}, \mathbf{l}, \mathbf{u}\|_\infty)^2)$. Finally, if $\hat{f}(\mathbf{x}_0) = 0$ then $\mathbf{x}_0$ is a feasible solution of (IP) and otherwise (IP) is infeasible. □

As a corollary of Lemmas 12 and 6, we immediately obtain that a polynomial (AugIP) algorithm is sufficient for solving (IP) in polynomial time:

Corollary 13 ((AugIP) $\implies$ (IP)). *Problem (IP) can be solved by solving $O(nL^2)$ instances of (AugIP), where $L := \log(\|A, f_{\max}, \mathbf{b}, \mathbf{l}, \mathbf{u}\|_\infty)$, plus time $O(n^\omega + \min\{n,m\}nm)$.*

3.2 Bounding Norms

We begin by using the Steinitz Lemma to obtain a basic bound on $g_1(A)$.

Proposition 14 (Steinitz [11], Sevastjanov, Banaszczyk [10]). *Let $\|\cdot\|$ be any norm, and let $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^d$ be such that $\|\mathbf{x}_i\| \le 1$ for $i \in [n]$ and $\sum_{i=1}^n \mathbf{x}_i = \mathbf{0}$. Then there exists a permutation $\pi \in S_n$ such that for each $k \in [n]$, $\|\sum_{i=1}^k \mathbf{x}_{\pi(i)}\| \le d$.*

Proof. We will inductively construct sets $A_n \supset A_{n-1} \supset A_{n-2} \supset \cdots A_m$ where $A_n = [n]$, $|A_k| = k$, and $\{\pi(i)\} = A_i \setminus A_{i-1}$, that is, $\pi(i)$ is the index which is removed as we go from A_i to A_{i-1}. (This also means we construct the permutation "backwards", first defining $\pi(n)$, then $\pi(n-1)$ etc. We stop at $\pi(n-m+1)$ because the sum of m vectors of norm at most 1 has norm at most m, regardless of the order.) The key is the following LP in variables $\boldsymbol{\lambda}$ (*not* $\mathbf{x}$; those are the input vectors, so they act as constants below):

$$\sum_{i \in A_k} \lambda_i \mathbf{x}_i = \mathbf{0}, \tag{3}$$

$$\sum_{i \in A_k} \lambda_i = k - m \tag{4}$$

$$\mathbf{0} \le \boldsymbol{\lambda} \le \mathbf{1} \tag{5}$$

Call it LP_k. First consider $k = n$. Observe that the LP is feasible: taking $\boldsymbol{\lambda} = \alpha\mathbf{1}$ satisfies (3) for any scalar $\alpha \in \mathbb{R}$, and choosing $\alpha = (n-m)/n$ scales the solution so that (4) is satisfied. Now assume that we are in step k of the construction, that is, we have a solution $\boldsymbol{\lambda}$ satisfying LP_k, and we wish to find an index to remove from A_k in order to construct A_{k-1} and define $\pi(k)$. A solution $\boldsymbol{\lambda}$ satisfying LP_k can be scaled down to satisfy $\sum_{i \in A_k} \lambda_i = k - m - 1$. Since there is a solution satisfying (3) and $\sum_{i \in A_k} \lambda_i = k - m - 1$, there is also a basic solution $\boldsymbol{\lambda}^*$ satisfying these constraints. Moreover, those are $m+1$ equality constraints and the rest are inequality constraints $\mathbf{0} \le \boldsymbol{\lambda} \le \mathbf{1}$. LP theory says that $\boldsymbol{\lambda}^*$ has at most $m+1$ fractional entries and the rest are integral, that is, either 0 or 1. We claim that at least one of the integer entries in $\boldsymbol{\lambda}^*$ is 0, and this is the index we will drop in order to define A_k. We have k variables in $\boldsymbol{\lambda}^*$, and they sum up to $k-m-1$. Say that there are $\varphi \le m+1$ fractional variables: they contribute strictly less than φ, say $\varphi - \epsilon$, with $\epsilon < m+1$. The remaining $k - \varphi$ integer variables must sum up to $k - m - 1 - (\varphi - \epsilon) = k - m - 1 - \varphi + \epsilon$, but this is impossible if they were all 1 because $\epsilon < m+1$ and $k - \varphi > k - m - 1 - \varphi + \epsilon$, equivalently, $k > k - m - 1 + \epsilon$. So there is an index i such that $\lambda_i^* = 0$ and we set $A_{k-1} = A_k \setminus \{i\}$.

Now it remains to verify that all prefix sums are indeed bounded by m. Consider:

$$\left\|\sum_{i=1}^k \mathbf{x}_{\pi(i)}\right\| = \left\|\sum_{i \in A_k} \mathbf{x}_i\right\| = \left\|\sum_{i \in A_k} (1-\lambda_i)\mathbf{x}_i\right\| \le \sum_{i \in A_k} (1-\lambda_i) = m$$

The first equality follows by the definition of π; the second follows by the fact that subtracting $\sum_i \lambda_i \mathbf{x}_i$ equates to subtracting $\mathbf{0}$ (by constraint (3)); the next inequality follows by assumption $\|\mathbf{x}_i\| \le 1$ for each i, and the final equality is by constraint (4). □

REFERENCES

[1] Daniel Dadush, Chris Peikert, and Santosh Vempala. 2011. Enumerative Lattice Algorithms in any Norm Via M-ellipsoid Coverings.. In *FOCS*, Rafail Ostrovsky (Ed.). IEEE, 580–589. http://dblp.uni-trier.de/db/conf/focs/focs2011.html#DadushPV11;http://dx.doi.org/10.1109/FOCS.2011.31;http://www.bibsonomy.org/bibtex/2eb00db06f2ab0db87494018ddc765f1e/dblp

[2] Jesús A. De Loera, Raymond Hemmecke, and Matthias Köppe. 2013. *Algebraic and Geometric Ideas in the Theory of Discrete Optimization.* MOS-SIAM Series on Optimization, Vol. 14. SIAM.

[3] Paul Gordan. 1873. Ueber die Auflösung linearer Gleichungen mit reellen Coefficienten. *Math. Ann.* 6, 1 (1873), 23–28.

[4] Jack E. Graver. 1975. On the foundations of linear and integer linear programming I. *Math. Program* 9, 1 (1975), 207–226.

[5] Martin Grötschel, László Lovász, and Alexander Schrijver. 1993. *Geometric algorithms and combinatorial optimization* (second ed.). Algorithms and Combinatorics, Vol. 2. Springer-Verlag, Berlin. xii+362 pages.

[8] Hendrik W. Lenstra, Jr. 1983. Integer programming with a fixed number of variables. *Mathematics of Operations Research* 8, 4 (1983), 538–548.

[9] Shmuel Onn. 2010. Nonlinear discrete optimization. *Zurich Lectures in Advanced Mathematics, European Mathematical Society* (2010). `http://ie.technion.ac.il/~onn/Book/NDO.pdf`.

[10] Sergey Sevast'janov and Wojciech Banaszczyk. 1997. To the Steinitz lemma in coordinate form. *Discrete Math.* 169, 1-3 (1997), 145–152.

[11] E. Steinitz. 1916. Bedingt konvergente Reihen und konvexe Systeme. *J. Reine Angew. Math.* 146 (1916), 1–52.

[12] Arne Storjohann and George Labahn. 1996. Asymptotically Fast Computation of Hermite Normal Forms of Integer Matrices. In *Proceedings of the 1996 International Symposium on Symbolic and Algebraic Computation, ISSAC '96, Zurich, Switzerland, July 24-26, 1996*, Erwin Engeler, B. F. Caviness, and Yagati N. Lakshman (Eds.). ACM, 259–266. http://dl.acm.org/citation.cfm?id=236869