Separable Convex Mixed-Integer Optimization: Improved Algorithms and Lower Bounds

Cornelius Brand
Chair of Algorithms and Complexity Theory, Faculty for Informatics and Computer Science, University of Regensburg, Germany

Martin Koutecký
Computer Science Institute, Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic

Alexandra Lassota
Eindhoven University of Technology, The Netherlands

Sebastian Ordyniak
University of Leeds, UK

Abstract

We provide several novel algorithms and lower bounds in central settings of mixed-integer (non-)linear optimization, shedding new light on classic results in the field. This includes an improvement on record running time bounds obtained from a slight extension of Lenstra's 1983 algorithm [Math. Oper. Res. '83] to optimizing under few constraints with small coefficients. This is important for ubiquitous tasks like knapsack–, subset sum– or scheduling problems [Eisenbrand and Weismantel, SODA'18, Jansen and Rohwedder, ITCS'19].

Further, we extend our algorithm to an intermediate linear optimization problem when the matrix has many rows that exhibit 2-stage stochastic structure, which adds to a prominent line of recent results on this and similarly restricted cases [Jansen et al. ICALP'19, Cslovjecsek et al. SODA'21, Brand et al. AAAI'21, Klein, Reuter SODA'22, Cslovjecsek et al. SODA'24]. We also show that the generalization of two fundamental classes of structured constraints from these works (n-fold and 2-stage stochastic programs) to separable-convex mixed-integer optimization are harder than their mixed-integer, linear counterparts. This counters a widespread belief popularized initially by an influential paper of Hochbaum and Shanthikumar, namely that "convex separable optimization is not much harder than linear optimization" [J. ACM '90].

To obtain our algorithms, we employ the mixed Graver basis introduced by Hemmecke [Math. Prog. '03], and our work is the first to give bounds on the norm of its elements. Importantly, we use these bounds differently from how purely-integer Graver bounds are exploited in related approaches, and prove that, surprisingly, this cannot be avoided.

2012 ACM Subject Classification Theory of computation → Convex optimization

Keywords and phrases Mixed-Integer Programming, Separable Convex Optimization, Parameterized Algorithms, Parameterized Complexity

Digital Object Identifier 10.4230/LIPIcs.ESA.2024.32

Related Version *Full Version*: `https://arxiv.org/abs/2111.08048v1`

Funding *Martin Koutecký*: Partially supported by Charles Univ. project UNCE 24/SCI/008 and by the project 22-22997S of GA ČR.
Sebastian Ordyniak: Supported by the Engineering and Physical Sciences Research Council (EPSRC, project EP/V00252X/1).

1 Introduction

We study the Mixed Integer Programming problem, which asks to minimize a (linear or non-linear) objective function over a linearly constrained, *mixed-integer* set of numerical variables, that is, some of them are required to be integral, while others may be fractional. The case where all variables are required to remain integral is the *purely-integer* case. This problem is of enormous importance for applications, as Bixby [3] says in his famous analysis of speed-ups for *linear* programming solvers: *"[I]nteger programming, and most particularly the mixed-integer variant, is the dominant application of linear programming in practice[,]"*. Moreover, concerning non-linear optimization, Bertsimas et al. note in their spectacular work [2] on the notorious subset selection problem in statistical learning that, over the past decades, algorithmic and hardware advances have elevated *convex* (and therefore, in particular, non-linear) mixed-integer optimization to a comparable level of relevance in applications.

Despite its practical ubiquity, the algorithmic theory of both linear and non-linear mixed integer optimization is much less developed compared to purely-integer problems. Indeed, the little insight we do have into the algorithmics of the mixed-integer case so far derives mainly from the purely-integer case (see the discussion of related work below): The best algorithms relevant to the setting of our article follow as a corollary of a 40 year-old result, namely, Lenstra's famous algorithm [23] (which can be extended, e.g. using [14, Theorem 6.7.9], to the setting of arbitrary convex target functions).

In this article, we focus on central special cases of the problem. We embark from the important case when there are only few constraints with small entries, which plays a key role in ubiquitous algorithmic applications such as scheduling problems, subset sum problems and knapsack-type problems, and has increasingly come into focus in recent years [13, 19]. The gained insight in this domain is then supercharged to attack a linear, mixed-integer optimization problem that arises in analogy to a long and prominent line of recent works on purely-integer problems with structured sets of constraints [10, 5, 9, 8, 12, 20].

Our Contributions

We design faster algorithms for non-linear mixed-integer optimization in the case where there are only few constraints, and the coefficients appearing in them are of small absolute value (Theorem 1). Building on this, we give parameterized algorithms for a generalized, hard *linear* mixed-integer optimization problem (Theorem 2). In addition, we prove new lower bounds for separable convex optimization problems, showing that the result on polynomial-time solvability of mixed-integer linear programming [4] subject to constraints of bounded treedepth does *not* translate to the separable convex case (Theorems 3 and 4). This may come as a surprise, considering the generally observed, tight connection between tractability for separable convex and linear target functions in the case of both fully continuous and purely-integer optimization (see [16, 6]), which might make it tempting to conjecture a similarly close relationship also in the mixed-integer case.

In terms of technical innovation, our algorithmic results are based on novel insights into as-of-yet poorly understood mixed Graver bases. We prove that, somewhat unexpectedly, the usual manner in which these bases are employed in the literature to aid the design of algorithms for integer programming problems can *not* yield much in the mixed-integer setting. One key contribution of our algorithmic results is that they demonstrate how to circumvent this limitation, and how to exploit the mixed Graver bases algorithmically nonetheless.

Importantly, and in contrast to most results in the mixed-integer realm to date, our results do *not* rely on expanding the results for purely-integral cases, which we prove to have only limited power, but instead, use new insights about the structure of the mixed-integer problem itself.

Results on Algorithms and Complexity of MIPs. To set the stage somewhat more formally, we consider with the following problem:

$$\min f(\mathbf{x}) : E\mathbf{x} = \mathbf{b}, \mathbf{l} \leq \mathbf{x} \leq \mathbf{u}, \mathbf{x} \in \mathbb{Z}^{n_\mathbb{Z}} \times \mathbb{R}^{n_\mathbb{R}}. \tag{MIP}$$

Here the number of columns is $n = n_\mathbb{Z} + n_\mathbb{R}$, the objective function $f : \mathbb{R}^n \to \mathbb{R}$ is *separable convex*, that is, $f(\mathbf{x}) = \sum_{i=1}^{n} f_i(x_i)$ for some sequence of *univariate* convex functions $f_1, \dots, f_n : \mathbb{R} \to \mathbb{R}$, $E \in \mathbb{Z}^{m \times n}$, E denotes the constraint matrix, $\mathbf{b} \in \mathbb{R}^m$ is the right-hand side, and the lower and upper bounds are $\mathbf{l}, \mathbf{u} \in (\mathbb{R} \cup \{\pm\infty\})^n$.

We set $\mathbb{X} = \mathbb{Z}^{n_\mathbb{Z}} \times \mathbb{R}^{n_\mathbb{R}}$, where $n_\mathbb{Z}$ and $n_\mathbb{R}$ should be clear from the context. An important special case is that of an integral right-hand side $\mathbf{b} \in \mathbb{Z}^m$ and bounds $\mathbf{l}, \mathbf{u} \in \mathbb{Z}^n$, and a linear target function $f(\mathbf{x}) = \mathbf{w}\mathbf{x} = \sum_i w_i x_i$. In this setting, (MIP) specializes to

$$\min \mathbf{w}\mathbf{x} : E\mathbf{x} = \mathbf{b}, \mathbf{l} \leq \mathbf{x} \leq \mathbf{u}, \quad \mathbf{x} \in \mathbb{X}. \tag{MILP}$$

On the front of algorithms for this problem, we improve the current, double-exponential record bound for mixed-integer programs with few rows and small coefficients to single-exponential, even when the target function is non-linear:[1]

▶ **Theorem 1** (Algorithm for MIPs with few rows)**.** *The problem* (MIP) *can be solved in single-exponential time* $(m\|E\|_\infty)^{\mathcal{O}(m^2)} \cdot \mathcal{R}$, *where* $\mathcal{R}$ *is the time needed to solve the continuous relaxation of any* (MIP) *with the constraint matrix* E.

Until now, the best way to solve a (MIP) with few rows and small coefficients would be to remove duplicate columns from E in a preprocessing step, and then use Lenstra's 1983 algorithm for mixed integer programming [23]. Since there are $2\|E\|_\infty + 1$ numbers of absolute value at most $\|E\|_\infty$, the preprocessing ensures that there are at most $(2\|E\|_\infty + 1)^m$ columns in E. This, however, leads to a *double-exponential* running time in terms of m.

Moreover, we use the above algorithm as a starting point for developing a novel algorithm for an intermediate problem. Namely, we now allow the bounds $\mathbf{l}, \mathbf{u} \in \mathbb{X}$ and right-hand side $\mathbf{b} \in \mathbb{R}^m$ to be fractional, that is, we consider the problem

$$\min \mathbf{w}\mathbf{x} : E\mathbf{x} = \mathbf{b}, \mathbf{l} \leq \mathbf{x} \leq \mathbf{u}, \quad \mathbf{x} \in \mathbb{X}. \tag{MILP$_\text{frac}$}$$

Already deciding feasibility of this variant has been shown to be NP-hard for totally unimodular matrices [7]. We are interested in algorithms that deal with constraint structures that were extensively treated in recent works in the purely integer setting [10, 5, 9, 8, 12, 20]. Namely, *n-fold* and *2-stage stochastic* matrices with bounded block-size, as depicted in Figure 1.[2] The matrices A_i and B_i in Figure 1 are called the *blocks* of the constraint matrices; furthermore, n denotes the number of blocks A_i and B_i. For the case of 2-stage stochastic constraints, we prove:

[1] As is common in the literature, we use the term *single-exponential* in x for functions of the form $2^{\text{poly}(x)}$, as opposed to e.g. $2^{O(x)}$. Similarly, we call exponential towers of height two, that is, $2^{2^{\text{poly}(x)}}$ *double-exponential* in x.

[2] Formal definitions of all terms used in the introduction will be given in the preliminaries.

$$\begin{pmatrix} B_1 & B_2 & \dots & B_n \\ A_1 & 0 & \dots & 0 \\ 0 & A_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A_n \end{pmatrix}$$

(a) An n-fold matrix

$$\begin{pmatrix} B_1 & A_1 & 0 & \dots & 0 \\ B_2 & 0 & A_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ B_n & 0 & 0 & \dots & A_n \end{pmatrix}$$

(b) A 2-stage stochastic matrix

Figure 1 The A_i and B_i are matrices of dimension bounded by a parameter. Note that n-folds and 2-stage stochastic matrices are transpositions of each other.

▶ **Theorem 2** (Algorithm for 2-stage stochastic (MILP$_{\text{frac}}$)). *The problem* (MILP$_{\text{frac}}$) *where E is a 2-stage stochastic matrix with block-dimensions $B_i \in \mathbb{Z}^{t\times r}$ and $A_i \in \mathbb{Z}^{t\times s}$ can be solved in time $g(r, s, \|E\|_\infty) \cdot n^r$, for some computable function g.*

Turning to lower bounds, we show that this result is likely optimal:

▶ **Theorem 3** (Hardness for 2-stage stochastic MIP). *The problem* (MIP) *with integral data is* W[1]*-hard when E is a 2-stage stochastic matrix with blocks of size bounded by a parameter and $\|E\|_\infty = 1$ already for linear objective functions.*

In particular, under the common parameterized complexity assumption that FPT $\neq$ W[1] holds, this rules out algorithms for (MIP) with running times of the form $g(k) \cdot \text{poly}(n)$, where k is the maximal block-dimension of each B_i, A_i in the 2-stage stochastic constraint matrices. Such a (double exponential) algorithm does exist for the pure integer case [12].

Moreover, we prove that the algorithm from Theorem 2 cannot be extended to the related case of n-fold constraint structure:

▶ **Theorem 4** (NP-hardness for n-fold MIP). *The problem* (MIP) *with integral data is* NP*-hard when E is an n-fold matrix with blocks of constant dimensions and $\|E\|_\infty = 1$ already for linear objective functions.*

Interestingly, the above hardness results demonstrate that the relationship between n-folds and 2-stage stochastic programs in the mixed case is different from purely-integer case: In the purely integer case, n-folds are solvable faster (in time FPT and single-exponentially [8]) than 2-stage stochastic programs [18], while in the mixed-integer case, the situation seems to be reversed.

Results on Mixed Graver Bases. Our algorithmic approach uses the mixed Graver basis of the constraint matrix. This is a mixed analogue of the usual integral Graver basis, which is a central object in all the recent developments around block-structured integer programs. Deeper insights into the Graver basis have led to new dynamic data structures [12], proximity theorems [8, 9, 12, 20, 21] and better convergence rate analyses [12]. Intuitively speaking, the elements of the Graver basis comprise all possible improving directions that have to be considered by an algorithm that seeks to iteratively augment suboptimal solutions.

The mixed Graver basis was introduced by Hemmecke [15] already in 2003, but not understood well enough to be used. On our way to showing Theorem 2, we prove several results about the mixed Graver basis which are of independent interest, and disprove the typical intuitions gained by studying the ordinary integral Graver basis. First, all elements of the *integral* Graver basis of an n-fold matrix with bounded block-dimension also have entries of bounded absolute value, whence they derive their algorithmic usefulness. We show that this is not true for the mixed Graver basis:

▶ **Theorem 5** (n-fold mixed Graver lower bound)**.** *There is an n-fold matrix E with constant-sized blocks and $\|E\|_\infty = 1$ such that the mixed Graver basis of E contains an element with 1-norm of size $\Omega(n)$.*

On the other hand, for 2-stage stochastic matrices, the ∞-norm of its elements can be bounded by a function of the block-dimensions and $\|E\|_\infty$:

▶ **Theorem 6** (2-stage stochastic mixed Graver upper bound)**.** *For any 2-stage stochastic matrix E, the maximum ∞-norm of an element of its mixed Graver basis is bounded by $h(r, s, \|E\|_\infty)$ for some computable function h.*

This bound also implies a proximity result: for any integer optimum $\mathbf{z}^*$, there is a nearby mixed optimum $\mathbf{x}^*$. Thus, we can first find $\mathbf{z}^*$ (which can be done efficiently), and then only search in a small neighborhood around $\mathbf{z}^*$.

Until now, a bound such as $h(r, s, \|E\|_\infty)$ on the Graver elements has always led to an algorithm with a corresponding running time $h(r, s, \|E\|_\infty)\,\mathrm{poly}(n)$. However, in the mixed case, such an algorithm is ruled out by Theorem 3. This shows that, in the mixed case, the common intuition of good bounds on the Graver norm directly leading to fast algorithms fails.

Related Work

We have already pointed to the most directly related recent works on block-structured (integer) linear programming. For an overview on the vast literature concerning practical attempts to deal with mixed-integer programming, the excellent article of Bertsimas et al. [2] provides pointers to relevant literature on this fascinating matter. We now sketch the theoretical literature in the field to contextualize the results obtained in the present paper, and in particular, how they contrast the typical, expected relationship between linear and convex optimization results observed heuristically in other situations. Since the limited insight we do have into the mixed-integer case so far derives mainly from the purely-integer case, we emphasize the comparison to the literature treating the latter setting to highlight patterns of lifting algorithmic results along two axes: From *purely-integer* to *mixed* domains, and from *linear* to *convex* objectives.

Towards the first axis, one first has to mention Lenstra's [23] algorithm for purely-integer *linear* optimization, which was seminal for the entire area. Notably, it extends to mixed-integer domains with a fixed number of integer coordinates. By the same token, there are other cases besides fixed integer dimension where tractability lifts from the purely-integer to the mixed case. For one, this includes the by-now classic polynomial-time solvability of linear integer optimization with totally unimodular constraint matrices [17]. In addition, it was shown more recently [4] how to obtain efficient algorithms for mixed-integer linear optimization subject to constraints of a particular structure (namely bounded treedepth, which includes n-fold and 2-stage stochastic programs), by leveraging the flurry of tractability results on purely-integer optimization in structurally restricted settings that we already pointed out above.

We now turn to the second axis, linear versus convex optimization. Indeed, as mentioned already above, the algorithm of Lenstra [23] for mixed- and purely-integer optimization with a fixed number of purely-integer coordinates can also be extended (e.g. using [14, Theorem 6.7.9]) to the setting of arbitrary convex target functions. In a similar vein, a general result of Hochbaum and Shanthikumar [16] establishes the following: Whenever the linear integer optimization problem with constraint matrices of bounded subdeterminants is polynomial-time solvable, then so is also the non-linear integer optimization problem for target functions that are *separable convex* (that is, a sum of univariate convex functions on the coordinates). Their paper's eponymous heuristic observation that "convex separable optimization is not much harder than linear optimization" has since become common wisdom, further consolidated e.g. by Chubanov's result on reducing linear to separable convex optimization over the fully (non-mixed) continuous domain. What is more, also the series of results on integer optimization subject to constraints of bounded treedepth mentioned above apply, especially, to separable convex target functions. It is worth noting that, since it is already NP-hard to optimize a *quadratic* (non-separable) convex function over the boolean hypercube $\{0,1\}^n$ [12, Proposition 101], and the constraint matrix describing (the facets of) the hypercube is both totally unimodular and of bounded treedepth, the algorithms for separable convex optimization are most likely not extensible to general convex objectives.

We emphasize that our results shed light on the algorithmic properties of mixed-integer optimization that defies the intuition that the body of work outlined above may suggest. Indeed, our results can be interpreted to mean that mixed-integer separable-convex optimization behaves rather unexpectedly from this point of view.

Organization

We give all necessary preliminaries in Sect. 2. Then, we give new results on mixed Graver bases and algorithmic consequences for mixed-integer linear programs with few rows in Sect. 3. In Sect. 4, we then extend this to an algorithm for the 2-stage stochastic case, and Sect. 5 contains a matching lower bound. In Sects. 6 and 7, we prove both complexity and Graver norm lower bounds for the n-fold case.

Due to the page limit, we postpone the proofs of some statements to the full version.

2 Preliminaries

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). Any (MIP) instance with infinite bounds $\mathbf{l}, \mathbf{u}$ can be reduced to an instance with finite bounds using standard techniques in polynomial time (solving the continuous relaxation and using proximity bounds to restrict the relevant region). So from now on we assume finite bounds $\mathbf{l}, \mathbf{u} \in \mathbb{X}$ with $\mathbb{X} = \mathbb{Z}^{n_\mathbb{Z}} \times \mathbb{R}^{n_\mathbb{R}}$

The set of indices at which $\mathbf{x}$ is non-zero is the *support* of $\mathbf{x}$, denoted $\mathrm{supp}(\mathbf{x})$. For positive integers $m \le n$ we set $[m, n] := \{m, \dots, n\}$ and $[n] := [1, n]$, and we extend this notation to 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 $\log := \log_2$. We define $\lfloor x \rceil$ to be $\lfloor x \rfloor$ if $x \ge 0$ and $\lceil x \rceil$ otherwise, and we define the *fractional part of* x to be $\{x\} := x - \lfloor x \rceil$. The division of variables into integer and continuous ones induces a division of the constraint matrix $E = (E_\mathbb{Z}\ E_\mathbb{R})$ where $E_\mathbb{Z} \in \mathbb{Z}^{m \times n_\mathbb{Z}}$ and $E_\mathbb{R} \in \mathbb{R}^{m \times n_\mathbb{R}}$, and analogously $\mathbf{x} = (\mathbf{x}_\mathbb{Z}, \mathbf{x}_\mathbb{R})$ and $f(\mathbf{x}) = f_\mathbb{Z}(\mathbf{x}_\mathbb{Z}) + f_\mathbb{R}(\mathbf{x}_\mathbb{R})$. More generally, whenever we make reference to any subset E' of columns or even submatrix of E, we will freely denote with $E'_\mathbb{Z}$ and $E'_\mathbb{R}$ the analogous division of E' into its integral and fractional part, respectively. Throughout, we assume that the rows of E are linearly independent.

We consider n-fold and 2-stage stochastic matrices. A matrix is of 2-stage stochastic structure if non-zero entries appear only in the first r columns and in n blocks of size $t \times s$ along the diagonal beside. The overall size is $nt \times (r + sn)$. An n-fold matrix is the transpose of a 2-stage stochastic matrix. It has thus $(r + sn)$ rows and nt columns. For an illustration, see Figure 1.

A vector $\mathbf{g} \in \ker(E) \setminus \{\mathbf{0}\}$ is a *circuit of* E if it is integral, its entries are co-prime, and it is support-minimal, that is, there is no vector $\mathbf{g}' \in \ker(E) \setminus \{\mathbf{0}\}$ with $\mathrm{supp}(\mathbf{g}') \subset \mathrm{supp}(\mathbf{g})$; let $\mathcal{C}(E)$ denote the set of circuits of E. For two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, we say that $\mathbf{x}$ *is conformal to* $\mathbf{y}$ and write $\mathbf{x} \sqsubseteq \mathbf{y}$ if, for each $i \in [n]$, $|x_i| \leq |y_i|$ and $x_i \cdot y_i \geq 0$. Intuitively, $\mathbf{x}$ and $\mathbf{y}$ are in the same orthant, and $\mathbf{y}$ is at least as far from $\mathbf{0}$ as $\mathbf{x}$ in each coordinate. We say that $\mathbf{x} = \sum_i \mathbf{g}_i$ is a *conformal sum* or a *conformal decomposition of* $\mathbf{x}$ if, for all i, $\mathbf{g}_i \sqsubseteq \mathbf{x}$. For an arbitrary set S, we write $\ker_S(E)$ as a shorthand for $\ker(E) \cap S$. In particular, the *mixed kernel of* E is defined as $\ker_{\mathbb{X}}(E)$. The *Graver basis of* E, denoted $\mathcal{G}(E)$, is the set $\mathcal{G}(E) = \{\mathbf{g} \in \ker_{\mathbb{Z}^n}(E) \setminus \{\mathbf{0}\} \mid \mathbf{g} \text{ is } \sqsubseteq\text{-minimal}\}$.

▶ **Definition 7** (Mixed Graver basis [15]). *Let* $E = (E_{\mathbb{Z}}\ E_{\mathbb{R}}) \in \mathbb{Z}^{m \times n}$*. The* mixed Graver basis $\mathcal{G}_{\mathbb{X}}(E)$ of E with respect to $\mathbb{X}$ *consists of all vectors* $(\mathbf{0}, \mathbf{g}_{\mathbb{R}})$*, where* $\mathbf{g}_{\mathbb{R}} \in \mathcal{C}(E_{\mathbb{R}})$*, together with all vectors* $(\mathbf{g}_{\mathbb{Z}}, \mathbf{g}_{\mathbb{R}}) \in \ker_{\mathbb{X}}(E)$ *such that* $\mathbf{g}_{\mathbb{Z}} \neq \mathbf{0}$ *and there is no* $(\mathbf{g}'_{\mathbb{Z}}, \mathbf{g}'_{\mathbb{R}}) \in (\ker_{\mathbb{X}}(E) \setminus \{\mathbf{0}\})$ *(unequal to* $(\mathbf{g}_{\mathbb{Z}}, \mathbf{g}_{\mathbb{R}})$*) such that* $(\mathbf{g}'_{\mathbb{Z}}, \mathbf{g}'_{\mathbb{R}}) \sqsubseteq (\mathbf{g}_{\mathbb{Z}}, \mathbf{g}_{\mathbb{R}})$*.*

For any p, $1 \leq p \leq \infty$, define $g_p^{\mathbb{X}}(E) := \max_{\mathbf{g} \in \mathcal{G}_{\mathbb{X}}(E)} \|\mathbf{g}\|_p$.

The following is a helpful trick to reduce a (MILP$_{\text{frac}}$) to a (MIP) with integer input data and a constraint matrix $(E\ I)$.

▶ **Lemma 8.** *Let an* (MILP$_{\text{frac}}$) *instance be given. It is possible to construct an equivalent* (MIP) *instance in linear time with a constraint matrix* $E' = (E\ I)$*, bounds* $\mathbf{l}', \mathbf{u}' \in \mathbb{Z}^{n+m}$*, and a right-hand side* $\mathbf{b}' \in \mathbb{Z}^m$*.*

Proof sketch. The proof works by first moving fractional right-hand sides into the lower and upper bounds. Then, we relax any fractional lower and upper bounds to their closest integers, and penalize violations of the original bounds in the new objective function, which is what introduces non-linearity (the resulting function is piece-wise linear convex with 3 pieces in each coordinate). ◀

3 The Basic Case: Matrices with Few Rows and Small Coefficients

This section develops the basic version of our algorithmic result. We begin by giving upper bounds for a certain notion of decompositions of elements in the mixed Graver basis, and then employ these bounds to our algorithmic ends.

3.1 Mixed-Graver Bound

We begin with an upper bound on the 1-norm for matrices with few rows and small coefficients. For this, we will need the Steinitz lemma:

▶ **Proposition 9** (Steinitz [26], Sevastjanov, Banaszczyk [25]). *Let* $\|\cdot\|$ *be any norm, and let* $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^d$ *be such that* $\|\mathbf{x}_i\| \leq 1$ *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* $k \in [n]$*,* $\|\sum_{i=1}^k \mathbf{x}_{\pi(i)}\| \leq d$*.*

▶ **Lemma 10.** *Let* $E \in \mathbb{Z}^{m \times (n_{\mathbb{Z}} + n_{\mathbb{R}})}$*. Then every* $\mathbf{g} \in \mathcal{G}_{\mathbb{X}}(E)$ *satisfies*

$$\|\mathbf{g}\|_1 \leq (2m\|E\|_\infty(2\|E\|_\infty + 1)^m + 1)^m + (2\|E\|_\infty + 1)^m$$

Proof. Let $\mathbf{g} \in \mathcal{G}_{\mathbb{X}}(E)$ and assume that all columns of E are distinct; we will show how to deal with doubled columns later. We define a sequence of vectors in the following manner: If $g_i \geq 0$, we add $\lfloor g_i \rfloor$ copies of the i-th column of E to the sequence, if $g_i < 0$ we add $|\lceil g_i \rceil|$ copies of the negation of column i to the sequence. Thus, for each $i \in [n]$, we obtained vectors $\mathbf{v}^i_1, \ldots, \mathbf{v}^i_{\lfloor g_i \rceil}$. Finally, we add the vector $\mathbf{o} = \sum_{i=1}^n \{g_i\} E_{\bullet,i}$ to the sequence. Notice that this vector is integral. Let q be the number of vectors in this sequence.

Clearly, the sequence of vectors sums up to $\mathbf{0}$ as it exactly corresponds to $E\mathbf{g}$ and $\mathbf{g} \in \ker_{\mathbb{X}}(E)$. Moreover, their ℓ_∞-norm is bounded by $\|E\|_\infty(2\|E\|_\infty + 1)^m$ since there are at most $(2\|E\|_\infty + 1)^m$ distinct columns, $\|E\|_\infty$ is the largest number appearing in any of them, and this is an upper bound on any number appearing in $\mathbf{o} = \sum_{i=1}^n \{g_i\} E_{\bullet,i}$. The remaining vectors $\mathbf{v}^i_j$ are bounded by $\|E\|_\infty$ in ℓ_∞-norm.

Using the Steinitz Lemma, there is a reordering $\mathbf{u}^1, \ldots, \mathbf{u}^q$ (i. e., $\mathbf{v}^i_j = \mathbf{u}^{\pi(i,j)}$ for some permutation π) of this sequence such that each prefix sum $\mathbf{p}_k := \sum_{j=1}^k \mathbf{u}^j$ is bounded by $m\|E\|_\infty(2\|E\|_\infty + 1)^m$ in the l_∞-norm. Clearly,

$$\left|\{\mathbf{x} \in \mathbb{Z}^m \mid \|\mathbf{x}\|_\infty \leq m\|E\|_\infty(2\|E\|_\infty + 1)^m\}\right| = (2m\|E\|_\infty(2\|E\|_\infty + 1)^m + 1)^m =: P.$$

Assume for contradiction that $q > P$. Then two of these prefix sums are the same, say, $\mathbf{p}_\alpha = \mathbf{p}_\beta$ with $1 \leq \alpha < \beta \leq q$. Obtain a vector $\mathbf{g}'$ from the sequence $\mathbf{u}^1, \ldots, \mathbf{u}^\alpha, \mathbf{u}^{\beta+1}, \ldots, \mathbf{u}^q$ as follows: begin with $g'_i := 0$ for each $i \in [n]$, and for every $\mathbf{u}^\ell$ in the sequence, set

$$g'_i := \begin{cases} g'_i + 1 & \text{if } \pi^{-1}(\ell) = (i,j) \text{ and } g_i \geq 0 \\ g'_i - 1 & \text{if } \pi^{-1}(\ell) = (i,j) \text{ and } g_i < 0 \\ g'_i + \{g_i\} & \text{if } \mathbf{u}^\ell = \mathbf{o}, \text{ for each } i \in [n] \ . \end{cases}$$

Here, (i,j) indicates the j-th copy of the i-th vector. Similarly obtain $\mathbf{g}''$ from the sequence $\mathbf{u}^{\alpha+1} \ldots, \mathbf{u}^\beta$. We have $E\mathbf{g}'' = \mathbf{0}$, as $\mathbf{p}_\alpha - \mathbf{p}_\beta = \mathbf{0}$ and thus, $\mathbf{g}'' \in \ker_{\mathbb{X}}(E)$ and hence, $\mathbf{g}' \in \ker_{\mathbb{X}}(E)$. Moreover, both $\mathbf{g}'$ and $\mathbf{g}''$ are non-zero and satisfy $\mathbf{g}', \mathbf{g}'' \sqsubseteq \mathbf{g}$. This is a contradiction with $\sqsubseteq$-minimality of $\mathbf{g}$ which is a condition needed for $\mathbf{g} \in \mathcal{G}_{\mathbb{X}}(E)$, hence $q \leq P$. Notice that only one of $\mathbf{g}'$ or $\mathbf{g}''$ may be fractional, as $\mathbf{o}$ will be in exactly one subsequence. For each of the at most $(2\|E\|_\infty + 1)^m$ columns, the respective fractional part in $\mathbf{g}$ contributes less than 1, so it follows that $\|\mathbf{g}\|_1 < P + (2\|E\|_\infty + 1)^m$ holds.

We are left to deal with the situation that E contains doubled columns. The solution is to adjust the construction of the sequence accordingly. Fix a column $E_{\bullet,i}$ and let S be the set of all indices j such that $E_{\bullet,i} = E_{\bullet,j}$. Let $u = \sum_{j \in S} g_j$. If $u > 0$, add $\lfloor u \rfloor$ copies of $E_{\bullet,i}$ into the sequence, else add $|\lceil u \rceil|$ copies of $-E_{\bullet,i}$ into the sequence. The contribution of this column type to $\mathbf{o}$ will be $\{u\}E_{\bullet,i}$. Since $-1 < \{u\} < 1$ for each column type, and the number of column types is bounded by $(2\|E\|_\infty + 1)^m$, our previous arguments hold. ◀

The proof of the above Lemma actually shows that there exists a particular decomposition of every element of $\ker_{\mathbb{X}}(E)$ into an element of $\ker_{\mathbb{Z}^n}(E)$ (which can be further decomposed into elements of $\mathcal{G}(E)$) and one element of $\ker_{\mathbb{X}}(E)$, which we can bound. This mixed element might not be an element of $\mathcal{G}_{\mathbb{X}}(E)$, and a bound on the elements of $\mathcal{G}_{\mathbb{X}}(E)$ does not imply a bound on this element. We crucially need this property in our proximity bound and the bounds on $\mathcal{G}_{\mathbb{X}}(E)$ for 2-stage matrices, as well as the prospect of extending these to multi-stage matrices. Thus, this emerges as an important feature:

▶ **Definition 11** (One-fat decomposition bound). *Let $\mathbf{x} \in \ker_{\mathbb{X}}(E)$. We say that $\mathbf{x} = \mathbf{h} + \mathbf{g}$ is a* one-fat decomposition *if it is a conformal decomposition, $\mathbf{h} \in \ker_{\mathbb{X}}(E)$ and $\mathbf{g} \in \ker_{\mathbb{Z}^n}(E)$, and we call $\mathbf{h}$ the* fat element *of the decomposition. For every p, $1 \leq p \leq \infty$, define $\mathbf{wt}^{\mathbb{X}}_p(\mathbf{x}) = \min \|\mathbf{h}\|_p$, where the minimum goes over all one-fat decompositions of $\mathbf{x}$. Define the* ℓ_p-weight of E with respect to $\mathbb{X}$ *as $\mathbf{wt}^{\mathbb{X}}_p(E) = \max_{\mathbf{x} \in \ker_{\mathbb{X}}(E)} \mathbf{wt}^{\mathbb{X}}_p(\mathbf{x})$.*

▶ **Corollary 12.** *For any matrix E, $\mathbf{wt}_1^{\mathbb{X}}(E) \le (2m\|E\|_\infty(2\|E\|_\infty + 1)^m + 1)^m$.*

Proof. Note that if $\mathbf{x} \in \ker_{\mathbb{X}}(E)$ is decomposable, then it has a decomposition into conformal $\mathbf{g}', \mathbf{g}''$, only one of which is fractional. Iterating this, we obtain the decomposition of $\mathbf{x}$ into several elements of $\mathcal{G}_{\mathbb{Z}^n}(E)$, and one element of $\ker_{\mathbb{X}}(E)$ which is bounded as stated. ◀

We will obtain a better bound on both $g_1^{\mathbb{X}}(E)$ and the ℓ_1-weight of E, using a recent result:

▶ **Proposition 13** ([24, Lemma 1]). *Let $\mathbf{x}^1, \dots, \mathbf{x}^n \in \mathbb{Z}^d$ and $\alpha_1, \dots, \alpha_n \in \mathbb{R}_+$ such that $\sum_{i=1}^n \alpha_i \mathbf{x}^i \in \mathbb{Z}^d$. If $\sum_{i=1}^n \alpha_i > d$, then there exist numbers $\beta_1, \dots, \beta_n \in \mathbb{R}_+$ such that, for all $i \in [n]$, $\beta_i \le \alpha_i$ and $\sum_{i=1}^n \beta_i \le d$, and $\sum_{i=1}^n \beta_i \mathbf{x}^i \in \mathbb{Z}^d$.*

An iterated use of this lemma gives rise to the following statement:

▶ **Lemma 14** (Packing Lemma). *Let $\mathbf{x}^1, \dots, \mathbf{x}^n \in \mathbb{Z}^d$ and $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n) \in \mathbb{R}_+^n$ such that $\sum_{i=1}^n \alpha_i \mathbf{x}^i \in \mathbb{Z}^d$. If $\sum_{i=1}^n \alpha_i > d$, there exist vectors $\boldsymbol{\beta}^1, \dots, \boldsymbol{\beta}^m \in \mathbb{R}_+^n$ such that, for each $j \in [m]$, $\boldsymbol{\beta}^j \le \boldsymbol{\alpha}$, $\sum_{i=1}^n \beta_i^j \mathbf{x}^i \in \mathbb{Z}^d$, $\|\boldsymbol{\beta}^j\|_1 \le d$, and $\sum_{j=1}^m \boldsymbol{\beta}^j = \boldsymbol{\alpha}$. Moreover, for all but at most one $j \in [m]$, $\|\boldsymbol{\beta}^j\|_1 \ge d/2$.*

Proof. The only potentially non-obvious part is the last sentence of the statement. Notice that if there are $\boldsymbol{\beta}^j$ and $\boldsymbol{\beta}^{j'}$, $j \neq j'$, with $\|\boldsymbol{\beta}^j\|_1, \|\boldsymbol{\beta}^{j'}\|_1 \le d/2$, then we can merge them. Formally, we set $\boldsymbol{\beta}^j := \boldsymbol{\beta}^j + \boldsymbol{\beta}^{j'}$, and delete $\boldsymbol{\beta}^{j'}$. ◀

Intuitively, the lemma allows us to take a non-negative linear combination of integer vectors whose result is an integer vector, and divide it into smaller such combinations while preserving the property that each smaller combination still results in an integer vector.

▶ **Lemma 15.** *Let $E \in \mathbb{Z}^{m \times (n_{\mathbb{Z}} + n_{\mathbb{R}})}$. Then $g_1^{\mathbb{X}}(E) \le (2m^2\|E\|_\infty + 1)^{m+1}$ and $\mathbf{wt}_1^{\mathbb{X}} \le (2m^2\|E\|_\infty + 1)^{2m+2}$.*

Proof sketch. As in Lemma 10, we will construct a sequence of vectors summing up to zero and then apply the Steinitz Lemma. However, this time we will use the Packing Lemma 14 to obtain a better bound on each element of the vector sequence and thus, a better bound on the elements of $\mathcal{G}_{\mathbb{X}}(E)$ overall. Unfortunately, this approach does not yield a one-fat decomposition, so we have to use the Steinitz Lemma in a more clever way to get a bound on $\mathbf{wt}_1^{\mathbb{X}}(E)$. ◀

The one-fat decomposition also allows us to prove a bound on the distance between an integer and mixed optimum, which we will use in both of our algorithmic results:

▶ **Lemma 16** (MIP Proximity). *Let $\mathbf{z}^* \in \mathbb{Z}^n$ be an integer optimum of a* (MIP) *instance, and let $\mathbf{x}^*$ be a mixed optimum closest to $\mathbf{z}^*$ in ℓ_p-norm, $1 \le p \le \infty$. Then $\|\mathbf{z}^* - \mathbf{x}^*\|_p \le \mathbf{wt}_p^{\mathbb{X}}(E)$.*

We will need a small technical proposition before we prove Lemma 16:

▶ **Proposition 17** ([12, Proposition 60]). *Let $\mathbf{x}, \mathbf{y}_1, \mathbf{y}_2 \in \mathbb{R}^n$, $\mathbf{y}_1, \mathbf{y}_2$ be from the same orthant, and f be a separable convex function. Then $f(\mathbf{x} + \mathbf{y}_1 + \mathbf{y}_2) - f(\mathbf{x} + \mathbf{y}_1) \ge f(\mathbf{x} + \mathbf{y}_2) - f(\mathbf{x})$.*

Proof of Lemma 16. Assume for contradiction that $\|\mathbf{z}^* - \mathbf{x}^*\|_p > \mathbf{wt}_p^{\mathbb{X}}(E)$. Since $(\mathbf{z}^* - \mathbf{x}^*) \in \ker_{\mathbb{X}}(E)$, it has a one-fat decomposition $\mathbf{h} + \mathbf{g}$ where $\|\mathbf{h}\|_p \le \mathbf{wt}_p^{\mathbb{X}}(E)$. As $\|\mathbf{z}^* - \mathbf{x}^*\|_p > \mathbf{wt}_p^{\mathbb{X}}(E)$, the integral part $\mathbf{g}$ is non-zero. Let $\hat{\mathbf{z}} := \mathbf{z}^* - \mathbf{g} = \mathbf{x}^* + \mathbf{h}$ and $\hat{\mathbf{x}} := \mathbf{x}^* + \mathbf{g} = \mathbf{z}^* - \mathbf{h}$. Thus, $\mathbf{z}^* - \mathbf{x}^* = \mathbf{h} + \mathbf{g} = (\mathbf{z}^* - \hat{\mathbf{x}}) + (\mathbf{z}^* - \hat{\mathbf{z}})$. Now Proposition 17 with $\mathbf{x} = \mathbf{x}^*$, $\mathbf{y}_1 = \mathbf{h}$, $\mathbf{y}_2 = \mathbf{g}$ shows

$$f(\mathbf{z}^*) - f(\hat{\mathbf{z}}) \ge f(\hat{\mathbf{x}}) - f(\mathbf{x}^*) \ .$$

By the conformality of the decomposition, $\hat{\mathbf{x}}$ and $\hat{\mathbf{z}}$ are within the $\mathbf{l}, \mathbf{u}$ bounds. As $\mathbf{g} \in \ker_{\mathbb{Z}^n}(E)$, $\hat{\mathbf{z}}$ is an integer feasible solution, and because $\mathbf{h} \in \ker_{\mathbb{X}}(E)$, $\hat{\mathbf{x}}$ is a mixed feasible solution. Furthermore, because $\mathbf{z}^*$ was an integer optimum and $\hat{\mathbf{z}}$ is integer feasible, the left hand side is non-positive, and so is $f(\hat{\mathbf{x}}) - f(\mathbf{x}^*)$, thus $\hat{\mathbf{x}}$ must be another mixed optimum and the right hand side must be zero, and so the left hand side, showing $\hat{\mathbf{z}}$ to be another integer optimum. However, $\hat{\mathbf{x}}$ is closer to $\mathbf{z}^*$, a contradiction. ◀

3.2 A Single-Exponential Algorithm

Armed with the bounds on the mixed Graver basis and our insights into one-fat decompositions, we are now ready to develop the single-exponential algorithm. Before we do so, however, a few general remarks are in order. These also apply to the two-stage stochastic algorithm for fixed block-dimensions later on.

▶ Remark 18. Both algorithmic results make use of the fact that if both the mixed and the integer version of the problem are feasible, then for every integral optimum, there is a mixed optimum nearby. It then suffices to first solve the (generally easier) integral version of the problem, and then solve an auxiliary mixed-integer program with the feasible region bounded by a small n-dimensional box around $\mathbf{x}$. Indeed, if $\mathbf{x}$ is an integral solution of $E\mathbf{x} = \mathbf{b}, \mathbf{l} \leq \mathbf{x} \leq \mathbf{u}$, then we will resort to solving the program $E(\mathbf{x} + \mathbf{y}) = \mathbf{b}$, $\|\mathbf{y} - \mathbf{x}\|_\infty \leq P$, $\mathbf{l} \leq \mathbf{x} + \mathbf{y} \leq \mathbf{u}$ for $\mathbf{y}$, which amounts to finding $\mathbf{y}$ with $E\mathbf{y} = \mathbf{0}$, $\mathbf{l}' \leq \mathbf{y} \leq \mathbf{u}'$ for some new bounds $\mathbf{l}', \mathbf{u}'$ such that $\|\mathbf{l}' - \mathbf{u}'\|_\infty$ is small. For general objectives, one optimizes the auxiliary objective $f'(\mathbf{y}) = f(\mathbf{x} + \mathbf{y})$, whereas for linear objectives no change is needed. Hence, all of the algorithmic heavy lifting will be done in order to solve problems of this form.

Of course, this strategy rests on the assumption that both the mixed and the integral variant of the problem are feasible. This assumption can in turn be removed by a standard two-phase approach, similar to what is customary e.g. for the Simplex algorithm, in order to find an initial feasible solution. In short, this is done by introducing slack variables that are penalized in the objective, but admit a trivial feasible solution. In the sequel, we will hence always assume feasibility.

We say that $\mathbf{x}_\epsilon$ is an *ϵ-accurate solution to* (MIP) if there exists an optimum $\mathbf{x}^*$ such that $\|\mathbf{x}^* - \mathbf{x}_\epsilon\|_\infty \leq \epsilon$. (For a discussion on the relationship of ϵ-accurate and ϵ-approximate optima and also the motivation to use the notion of ϵ-accuracy, see [16, Section 1.2].)

▶ **Theorem 1** (Algorithm for MIPs with few rows). *The problem* (MIP) *can be solved in single-exponential time* $(m\|E\|_\infty)^{\mathcal{O}(m^2)} \cdot \mathcal{R}$*, where* $\mathcal{R}$ *is the time needed to solve the continuous relaxation of any* (MIP) *with the constraint matrix* E*.*

Proof of Theorem 1. The integer problem can be solved in time $(m\|E\|_\infty)^{\mathcal{O}(m^2)} + \mathcal{R}(\epsilon)$ by known techniques [12, 13] where $\mathcal{R}(\epsilon)$ is the ϵ-accurate solution to the continuous relaxation – essentially, first solve the continuous relaxation, then reduce $\mathbf{b}, \mathbf{l}, \mathbf{u}$ using proximity bounds, then solve a dynamic program. Now by Lemma 16, a mixed optimum $\mathbf{x}^*$ is at most $\mathbf{wt}_1^{\mathbb{X}}(E) \leq (2m^2\|E\|_\infty + 1)^{2m+2} =: P$ far in 1-norm. The proximity bound implies that all prefix sums of $\mathbf{x}^*_{\mathbb{Z}}$ with $E_{\mathbb{Z}}$ belong to the integer box $R := [-\|E\|_\infty \cdot P, \|E\|_\infty \cdot P]^m$, which has at most $(2\|E\|_\infty \cdot P + 1)^m = (m\|E\|_\infty)^{\mathcal{O}(m^2)}$ elements.

This allows us to construct a dynamic program with $n_{\mathbb{Z}} + 1$ stages. Our DP table D shall have an entry $D(i, \mathbf{r})$ for $i \in [n_{\mathbb{Z}}]$ and $\mathbf{r} \in R$ whose meaning is the minimum objective attainable if the prefix sum of $\mathbf{x}^*_{\mathbb{Z}}$ and $E_{\mathbb{Z}}$ restricted to the first i coordinates is $\mathbf{r}$. To that end, for all $\mathbf{r} \in R$, define $x_i^*(\mathbf{r})$ to be the choice of $x_i^* \in [-P, P]$ which minimizes f_i and such that $E_{\bullet,i} x_i^* = \mathbf{r}$; it is possible for the solution to be undefined if no number in $[-P, P]$

satisfies the conditions. Similarly, define $\mathbf{x}^*_{\mathbb{R}}(\mathbf{r})$ to be an ϵ-accurate minimizer of $f_{\mathbb{R}}$ satisfying $E_{\mathbb{R}}\mathbf{x}^*_{\mathbb{R}} = \mathbf{r}$. To compute D, set $D(0,\mathbf{r}) := 0$ for $\mathbf{r} = \mathbf{0}$ and $D(0,\mathbf{r}) := +\infty$ otherwise, and for $i \in [n_{\mathbb{Z}}]$, set

$$D(i,\mathbf{r}) := \min_{\substack{\mathbf{r}',\mathbf{r}''\in R:\\ \mathbf{r}'+\mathbf{r}''=\mathbf{r}}} D(i-1,\mathbf{r}') + f^i(x^*_i(\mathbf{r}'')) \enspace .$$

The last stage is defined as

$$D(n_{\mathbb{Z}}+1,\mathbf{0}) := \min_{\substack{\mathbf{r}',\mathbf{r}''\in R:\\ \mathbf{r}'+\mathbf{r}''=\mathbf{0}}} D(n_{\mathbb{Z}},\mathbf{r}') + f_{\mathbb{R}}(\mathbf{x}^*_{\mathbb{R}}(\mathbf{r}'')) \enspace .$$

The value of the optimal solution is $D(n_{\mathbb{Z}}+1,\mathbf{0})$ and the solution $\mathbf{x}^*$ itself can be computed easily with a bit more bookkeeping in the table D.

As for complexity, the first $n_{\mathbb{Z}}$ stages of the DP can be computed in time at most $n_{\mathbb{Z}} \cdot |R|^2 = (m\|E\|_\infty)^{\mathcal{O}(m^2)} n_{\mathbb{Z}}$, and the last stage solves the continuous relaxation $|R|$ times, taking time $|R|\mathcal{R}(\epsilon)$. Altogether, the algorithm takes time at most $(m\|E\|_\infty)^{\mathcal{O}(m^2)}\mathcal{R}(\epsilon)$. Regarding correctness, note that any ϵ-accurate solution $\mathbf{x}^*$ is such that $\mathbf{x}^*_{\mathbb{R}}$ is an ϵ-accurate minimizer of $E_{\mathbb{R}}\mathbf{x}_{\mathbb{R}} = -E_{\mathbb{Z}}\mathbf{x}^*_{\mathbb{Z}}$, $\mathbf{l}_{\mathbb{R}} \le \mathbf{x}_{\mathbb{R}} \le \mathbf{u}_{\mathbb{R}}$, and $\mathbf{x}^*_{\mathbb{Z}}$ is an integer minimizer of $E_{\mathbb{Z}}\mathbf{x}_{\mathbb{Z}} = -E_{\mathbb{R}}\mathbf{x}^*_{\mathbb{R}}$, $\mathbf{l}_{\mathbb{Z}} \le \mathbf{x}_{\mathbb{Z}} \le \mathbf{u}_{\mathbb{Z}}$. Since the algorithm finds exactly such minimizers, its correctness follows. ◀

4 Algorithms for the 2-Stage Stochastic Case

After giving the basic version of our algorithm for the case of few rows, we now develop our algorithm for the case of fixed block-dimension. We first prove our bound of the mixed Graver basis:

▶ **Theorem 6** (2-stage stochastic mixed Graver upper bound). *For any 2-stage stochastic matrix E, the maximum ∞-norm of an element of its mixed Graver basis is bounded by $h(r,s,\|E\|_\infty)$ for some computable function h.*

Proof sketch. We first decompose the element of $\mathcal{G}_{\mathbb{X}}(E)$ blockwise according to Lemma 15. Then, we apply a recent result of [9] on the existence of submultisets with equal sums in certain multisets of vectors with similar (but, notably, not identical) sums. This is made possibly by the fact that we have not only a bound on $g^{\mathbb{X}}_\infty(E)$ but a bound on the weight of a one-fat decomposition. This gives us the desired one-fat decomposition for the 2-stage stochastic case. ◀

From Theorem 6 and Lemma 16, it follows that:

▶ **Corollary 19.** *Let $\mathbf{z}^* \in \mathbb{Z}^n$ be an integer optimum of a 2-stage stochastic* (MIP) *instance. Then there exists a mixed optimum $\mathbf{x}^* \in \mathbb{X}$ such that $\|\mathbf{z}^* - \mathbf{x}^*\|_\infty \le h(r,s,\|E\|_\infty)$ for a double exponential function h.*

4.1 A Polynomial Algorithm for Fixed Block-Dimension

Using the upper bounds for 2-stage stochastic MIPs on proximity and weight as combined in Corollary 19, we can now formulate an algorithm which solves the 2-stage stochastic MILP problem in polynomial time whenever the block-dimensions are fixed. We recall that h is the function from Theorem 6. In accordance with Remark 18, we note two things: Firstly, by following a standard two-phase approach, we may assume that the problem at hand is integrally feasible. Then, secondly, the algorithm solves the integer program corresponding to the instance to optimality, which is fixed-parameter tractable [1, 22]. We thereby obtain an integer optimum $\mathbf{z}^*$, and we can now restrict ourselves to solving the following auxiliary MILP to optimality:

$$\min \mathbf{w}\mathbf{x}: \ E\mathbf{x} = \mathbf{0}, \hat{\mathbf{l}} \leq \mathbf{x} \leq \hat{\mathbf{u}}, \mathbf{x} \in \mathbb{Z}^{n_\mathbb{Z}} \times \mathbb{R}^{n_\mathbb{R}}, \mathbf{l}, \mathbf{u} \in \mathbb{R}^n, \mathbf{b} \in \mathbb{Z}^m. \tag{AuxMILP}$$

Here, $\hat{\ell}_i = \max\{\ell_i - z_i^*, -h(r, s, \|E\|_\infty)\}$ and $\hat{u}_i = \min\{u_i - z_i^*, h(r, s, \|E\|_\infty)\}$. Observe that

$$\|\hat{\mathbf{l}} - \hat{\mathbf{u}}\|_\infty \leq 2h(r, s, \|E\|_\infty) \tag{1}$$

holds. For an optimal solution $\mathbf{x}^*$ to (AuxMILP), the augmented solution $\mathbf{x}^* + \mathbf{z}^*$ is then an optimal solution to the original MILP, by Corollary 19.

What remains is to show how to solve (AuxMILP) in the claimed time bound. This is effected by proving the following Lemma:

▶ **Lemma 20.** *Let V be the set of vertices of all integer slices of the auxiliary mixed-integer program* (AuxMILP)*. There are at most $(8h(r, s, \|E\|_\infty))^{(r+1)(s+1)} n^r$ distinct global parts appearing in V, and they can be enumerated with polynomial delay.*

Proof sketch. The proof goes by analyzing the structure of invertible submatrices of two-stage stochastic matrices. Then, it becomes apparent that the global part of a basic solution is essentially determined by which subset of r blocks out of all n blocks influences the global part. The number of such choices is clearly bounded by n^r. The remainder of the bound stems from various guessing steps, including some of the values for the integer variables. Hence the appearance of h in the bound, making also the bounds from Theorem 6 crucially come into play. ◀

Lemma 20 now suggests an obvious strategy to solve the (AuxMILP) to optimality:

▶ **Proposition 21.** (AuxMILP) *can be solved in time $h(r, s, \|E\|_\infty)^{O(rs)} \cdot n^r$.*

Proof sketch. By Lemma 20, we may enumerate all possible global parts of vertices in the required time bound, guess the corresponding global integer values, and then solve the resulting block-diagonal mixed-integer system to optimality using the algorithm of Theorem 1 (notice that here we are in the special case of LP which can be solved exactly, i.e., with $\epsilon = 0$, and in strongly polynomial time since $\|E\|_\infty$ is small, so $\mathcal{R}(0) = \text{poly}(n)$). Among all choices of global parts, pick the one that yields the optimal value for the full program. ◀

We have now obtained:

▶ **Theorem 2** (Algorithm for 2-stage stochastic $(\text{MILP}_{\text{frac}})$)**.** *The problem* $(\text{MILP}_{\text{frac}})$ *where E is a 2-stage stochastic matrix with block-dimensions $B_i \in \mathbb{Z}^{t \times r}$ and $A_i \in \mathbb{Z}^{t \times s}$ can be solved in time $g(r, s, \|E\|_\infty) \cdot n^r$, for some computable function g.*

Proof. As mentioned before, it is enough to first solve the integer program corresponding to the MILP instances, and then solving the auxiliary problem using Proposition 21. ◀

▶ Remark 22. Let us note two things: Firstly, the exponent of n in our algorithm is only dependent on the number r of global variables. Hence, for values of s such that $h(r, s, \|E\|_\infty)^s \leq n^{f(r)}$ for some function f, our algorithm remains polynomial for fixed r.

Secondly, note that we may choose strongly polynomial (or rather, strongly fpt) subroutines to solve the arising integer and mixed-integer programs. In this case, also the algorithm we obtain is strongly polynomial for fixed block-dimensions.

5 W[1]-Hardness of 2-Stage Stochastic MILPs with Fractional Bounds

In the following we show that 2-stage stochastic (MILP$_{\text{frac}}$) and (MIP) with integral data is W[1]-hard parameterized by the block-dimension even if $\|E\|_\infty = 1$.

▶ **Theorem 3** (Hardness for 2-stage stochastic MIP). *The problem* (MIP) *with integral data is* W[1]*-hard when E is a 2-stage stochastic matrix with blocks of size bounded by a parameter and $\|E\|_\infty = 1$ already for linear objective functions.*

Proof Sketch of Theorem 3. We show the theorem using a parameterized reduction from the well-known Subset Sum problem, which is W[1]-hard when parameterized by the number of elements in a solution [11].

Subset Sum
Input: A set A of pairwise distinct natural numbers and two natural numbers k and t.
Goal: Decide whether there is a subset $S \subseteq A$ with $|S| = k$ and $\sum_{s \in S} s = t$?

Transformation: We give a formulation of Subset Sum as a 2-stage stochastic MILP. To do so, we first scale all input numbers $a_1, a_2, \ldots, a_n$ in A and t by $1/\max_i\{a_i\}$. Denote the new numbers as $a'_1, a'_2, \ldots, a'_n$ and t'. The scaling ensures that all considered sums are smaller or equal to 1, which comes in handy later on.

Let x^i_j be a binary variable that will indicate that a'_i is the jth number appearing in the sum for all $i \in [n]$ and $j \in [k]$. We collect those numbers not appearing in a solution in a binary slack variable x^i_{k+1} for each $i \in [n]$, yielding the constraints:

$$\sum_{j=1}^{k+1} x^i_j = 1 \qquad \forall i \in [n] \tag{2}$$

To express the condition on the sum of the solution being t', we introduce fractional variables y^i_j that take on the value a'_i if and only if $x^i_j = 1$ for $i \in [n]$ and $j \in [k]$. While this is trivially achieved by $y^i_j = a'_i x^i_j$, the crux is to model this without including a'_i as a coefficient, which would not be bounded by the parameter any more. This is accomplished by requiring the following:

$$y^i_j \le x^i_j \qquad \forall i \in [n], \forall j \in [k] \tag{3}$$

$$\sum_{j=1}^{k+1} y^i_j = a'_i \qquad \forall i \in [n] \tag{4}$$

This has the intended effect since $a'_i \le 1$ by construction. We will then store the solution indicated by the assignment to the x^i_j variables in yet another set of variables, denoted as z_j, where j ranges from 1 to k

$$\sum_{j=1}^{k} z_j = t' \tag{5}$$

While it is easy to project the y^i_j to z_j, the straightforward way to do so would blow up the block size to $\Omega(n)$. Indeed, to ensure that the z_j have the intended semantics, consider the following: The equality $z_j = y^i_j$ ought to be satisfied for exactly one choice of i, say when $i = i'$ (assuming distinct inputs); otherwise, $z_j = y^i_j + s^i_j$ holds for some non-zero compensation term s^i_j, whenever $i \neq i'$. Note that, while the s^i_j do satisfy a function similar to slack variables, they may well need to be negative. In addition, we introduce binary variables r^i_j for all $i \in [n]$ and $j \in [k]$, indicating whether or not $s^i_j = 0$. The above semantics are captured in the following constraints:

$$z_j = y^i_j + s^i_j \qquad \forall i \in [n], \forall j \in [k] \tag{6}$$

$$z_j \geq \min_i a'_i \tag{7}$$

$$-r^i_j \leq s^i_j \leq r^i_j \qquad \forall i \in [n], \forall j \in [k] \tag{8}$$

Our aim is then to minimize the number of times any of the s^i_j are used, or conversely, to make $z_j = y^i_j$ for some i as often as possible, which is expressed in the choice of the objective function

$$\min \sum_{j=1}^{k} \sum_{i=1}^{n} r^i_j \tag{9}$$

As argued, note that in a solution of a yes instance, for a fixed j, $z_j = y^i_j \geq \min_i a'_i$ (equivalently, $r^i_j = 0$) holds for exactly one choice of i, making the optimum equal to $k(n-1)$.

The above constraints define a 2-stage stochastic MILP formulation with fractional variables z_j, y^i_j and s^i_j, and binary variables x^i_j and r^i_j. The global part is made up by the z_j, of which there are k. The remaining variables are distributed across n blocks of dimension $O(k)$ each, including the respective slack variables for the inequality constraints. The largest entry in the constraint matrix is $1 = O(k)$, and clearly, the transformation can be carried out in time polynomial in n and k. ◀

6 NP-hardness of n-Fold MIPs

The algorithmic upper bound for 2-stage stochastic programs stands in contrast to a much stronger bound for the n-fold case. Namely, we show NP-hardness of n-fold (MILP$_{\text{frac}}$) for constant parameter values. By Lemma 8, we immediately get that n-fold (MIP) is also NP-hard for constant parameter values.

▶ **Theorem 4** (NP-hardness for n-fold MIP). *The problem* (MIP) *with integral data is* NP*-hard when E is an n-fold matrix with blocks of constant dimensions and $\|E\|_\infty = 1$ already for linear objective functions.*

Proof of Theorem 4. We reduce from the well-known Partition problem. That is, given integers $a_1, \ldots, a_n$ the Partition problems asks for the existence of a subset $I \subseteq [n]$ such that $\sum_{i \in I} a_i = \sum_{i \notin I} a_i$.

Let an instance of Partition be given. Without loss of generality, assume that $a_{\max} := \max_i a_i \leq 1$; this can be achieved, e. g., by scaling every number of the original instance by $1/a_{\max}$. We will have n bricks, with brick $i \in [n]$ representing the choice whether $i \in I$ or $i \notin I$.

Specifically, for each $i \in [n]$, introduce integer variables $x^i_1, x^i_2 \in \{0, 1\}$ and continuous variables y^i_1, y^i_2 with bounds $0 \leq y^i_1, y^i_2 \leq 1$. The local constraints (matrix A) are as follows. We enforce a disjunction on the x-variables by the constraint $x^i_1 + x^i_2 = 1$ for every i and we enforce that $y^i_1 = a_i$ iff $x^i_1 = 1$ and similarly $y^i_2 = a_i$ iff $x^i_2 = 1$ by the constraints $y^i_1 + y^i_2 = a_1$, $y^i_1 \leq x^i_1$, and $y^i_2 \leq x^i_2$ for every i.

It is now easy to see that the following global constraint encodes the requirement that $\sum_{i \in I} a_i = \sum_{i \notin I} a_i$:

$$\sum_{i=1}^{n} y_1^i = \sum_{i=1}^{n} y_2^i. \tag{10}$$

Altogether, the instance has four variables per block, four local constraints and one global constraint, and is feasible if and only if the original PARTITION instance is. ◀

7 Lower Bound on the Graver Norm of n-fold MIPs

In this section, we show that the 1-norm of the mixed Graver norm can be unbounded even for n-fold matrices.

We start with the following auxiliary lemma, which is crucial for constructing an element of the mixed Graver basis with unbounded 1-norm.

▶ **Lemma 23.** *Let n be an integer. There are two sets S and T of natural numbers with $|S| = |T| = n$ such that:*

(1) $\sum_{s \in S} s = \sum_{t \in T} t = 2^{n^2} - 1$ *and*

(2) *for every two subsets $S' \subseteq S$ and $T' \subseteq T$, with $0 < |S' \cup T'| < 2n$, it holds that $\sum_{s \in S'} s \neq \sum_{t \in T'} t$.*

Proof. Let $X \subseteq \mathbb{N} \setminus \{0\}$. We denote by $N(X)$, the natural number whose binary representation has a 1 at the i-th bit (with 1 being the lowest-value bit) if and only if $i \in X$. Conversely, for a natural number x, let $B(x)$ be the set of all indices i such that the binary representation of x is 1 at the i-th bit. Note that $B(N(X)) = X$ for every $X \subseteq \mathbb{N} \setminus \{0\}$.

For every i and j with $1 \leq i, j \leq n$, let $p(i, j) = (i - 1)n + j$. For every i with $1 \leq i \leq n$, we set:

- s_i is equal to $N(R_i)$, where $R_i = \{p(i, j) \mid 1 \leq j \leq n\}$,
- t_i is equal to $N(C_i)$, where $C_i = \{p(j, i) \mid 1 \leq j \leq n\}$.

We claim that setting $S = \{s_1, \ldots, s_n\}$ and $T = \{t_1, \ldots, t_n\}$ satisfies the statement of the lemma: As $\{B(s_1), \ldots, B(s_n)\}$ and $\{B(t_1), \ldots, B(t_n)\}$ form a partition of $[n^2]$, it holds that $\sum_{s \in S'} s = N(\bigcup_{s \in S'} B(s))$ and $\sum_{t \in T'} t = N(\bigcup_{t \in T'} B(t))$ for every subsets $S' \subseteq S$ and $T' \subseteq T$. Therefore, $\sum_{s \in S} s = \sum_{t \in T} t = N([n^2]) = 2^{n^2} - 1$, which shows (1).

Towards showing (2), let S' and T' be any two subsets with $S' \subseteq S$ and $T' \subseteq T$ such that $0 < |S' \cup T'| < 2n$. As $0 < |S' \cup T'| < 2n$, we obtain that either:

- there are i and j with $1 \leq i, j \leq n$ such that $s_i \in S \setminus S'$ and $t_j \in T'$ or
- there are i and j with $1 \leq i, j \leq n$ such that $t_i \in T \setminus T'$ and $s_j \in S'$.

Since the proofs for the two cases are analogous, we only give the proof for the former case. Let $O = B(s_i) \cap B(t_j)$ and note that $O = R_i \cap C_j = \{p(i, j)\} \neq \emptyset$. Since $t_i \in T'$, it holds that $O \in \bigcup_{t \in T'} B(t)$. However, due to $s_i \notin S'$, we have that $O \notin \bigcup_{s \in S'} B(s)$. Consequently, $\bigcup_{s \in S'} B(s) \neq \bigcup_{t \in T'} B(t)$ and therefore also $\sum_{s \in S'} s \neq \sum_{t \in T'} t$. ◀

▶ **Theorem 5** (n-fold mixed Graver lower bound)**.** *There is an n-fold matrix E with constant-sized blocks and $\|E\|_\infty = 1$ such that the mixed Graver basis of E contains an element with 1-norm of size $\Omega(n)$.*

Proof sketch. Let n be an integer, $\mathbb{X}_n = (\mathbb{Z} \times \mathbb{R} \times \mathbb{R})^n$, and E_n be the matrix given by the n-fold of $\begin{pmatrix} 0 & I_3 \\ 0 & A \end{pmatrix}$, where I_3 is the identity matrix of dimension 3 and $A = (1, 1, 1)$. Note first that the structure of the matrix E together with the fact that the first coordinate in each block is integer ensures that any vector of the form $(-1, s/V, 1-s/V)$ and $(1, -t/V, 1+t/V)$ for some s, t, V with $0 \le s, t \le V$ is contained in the mixed Graver basis $\mathcal{G}_{\mathbb{X}_1}(E)$ of E. This allows us to construct $\mathbf{g}$ by using $n/2$ blocks of the form $(-1, s/V, 1-s/V)$ and $n/2$ blocks of the form $(1, -t/V, 1+t/V)$, where $s \in S$ and $t \in T$ for some well-constructed sets S and T of integers. Moreover, because of the first three rows (given by n-repetitions of I_3) the sum of the i-th coordinate over all blocks has to be 0. Therefore, to force that all n blocks of $\mathbf{g}$ use non-zero kernel elements (of $\mathcal{G}_{\mathbb{X}_1}(E)$), it suffices to construct the sets S and T in such a way that $\sum_{s \in S'} s = \sum_{t \in T'} t$ for some subsets $S' \subseteq S$ and $T' \subseteq T$ if and only if $S' = S$ and $T' = T$. We show that this is possible in an auxiliary lemma. ◀

8 Open Questions

Our work points towards two main directions for further research. First, note that we formulate our algorithms in reference to 2-stage stochastic constraint matrices. As mentioned, these are generalized by matrices of bounded primal treedepth, so-called multi-stage stochastic matrices. They are structured in much the same way as in Fig. 1, but with diagonal blocks of recursive multi-stage stochastic form (and the depth of this recursion is bounded). Judging from previous results in the area, there is reason to believe that our algorithmic results generalize to multi-stage stochastic programs. However, some caution seems appropriate. After all, one key takeaway of both the lower bounds and the algorithms shown in this paper is that block-structured mixed-integer programs do not behave as predictably as one might hope.

A second natural, much more ambitious direction of investigation is to try to extend the present algorithmic results on linear optimization to arbitrary separable convex objective functions. Despite significant efforts, we were not able to push beyond the algorithms obtained here. As for some intuition on why extending Theorem 2 to the separable convex case in the style of Theorem 1 seems to fail: For the latter, the search space for optima is naturally restricted already by the fact that there are only few constraints. For the former, however, such a restriction is not possible based on rows alone; instead, we argue about vertices of the associated polytope, which are related to mixed-integer optimal solutions. In the separable convex case, this connection between optimal solutions and vertices disappears, and we are left with no handle on the size of the search space. We consider the problem of circumventing these roadblocks an intriguing and hard open question.

References

1. Matthias Aschenbrenner and Raymond Hemmecke. Finiteness theorems in stochastic integer programming. *Foundations of Computational Mathematics*, 7(2):183–227, 2007.
2. Dimitris Bertsimas, Angela King, and Rahul Mazumder. Best subset selection via a modern optimization lens. *The Annals of Statistics*, 44(2):813–852, 2016. URL: `http://www.jstor.org/stable/43818629`.
3. Robert E Bixby. Solving real-world linear programs: A decade and more of progress. *Operations research*, 50(1):3–15, 2002.
4. Cornelius Brand, Martin Koutecký, and Sebastian Ordyniak. Parameterized algorithms for milps with small treedepth. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*, pages 12249–12257. AAAI Press, 2021. URL: `https://ojs.aaai.org/index.php/AAAI/article/view/17454`, `doi:10.1609/AAAI.V35I14.17454`.

5 Timothy F. N. Chan, Jacob W. Cooper, Martin Koutecký, Daniel Král, and Kristýna Pekárková. Matrices of optimal tree-depth and a row-invariant parameterized algorithm for integer programming. *SIAM J. Comput.*, 51(3):664–700, 2022. `doi:10.1137/20M1353502`.

6 Sergei Chubanov. A polynomial-time descent method for separable convex optimization problems with linear constraints. *SIAM J. Optim.*, 26(1):856–889, 2016. `doi:10.1137/14098524X`.

7 Michele Conforti, Marco Di Summa, Friedrich Eisenbrand, and Laurence A. Wolsey. Network formulations of mixed-integer programs. *Math. Oper. Res.*, 34(1):194–209, 2009. `doi:10.1287/moor.1080.0354`.

8 Jana Cslovjecsek, Friedrich Eisenbrand, Christoph Hunkenschröder, Lars Rohwedder, and Robert Weismantel. Block-structured integer and linear programming in strongly polynomial and near linear time. In Dániel Marx, editor, *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms, SODA 2021, Virtual Conference, January 10 - 13, 2021*, pages 1666–1681. SIAM, 2021. `doi:10.1137/1.9781611976465.101`.

9 Jana Cslovjecsek, Friedrich Eisenbrand, Michal Pilipczuk, Moritz Venzin, and Robert Weismantel. Efficient sequential and parallel algorithms for multistage stochastic integer programming using proximity. In Petra Mutzel, Rasmus Pagh, and Grzegorz Herman, editors, *29th Annual European Symposium on Algorithms, ESA 2021, September 6-8, 2021, Lisbon, Portugal (Virtual Conference)*, volume 204 of *LIPIcs*, pages 33:1–33:14. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021. `doi:10.4230/LIPIcs.ESA.2021.33`.

10 Jana Cslovjecsek, Martin Koutecky, Alexandra Lassota, Michal Pilipczuk, and Adam Polak. Parameterized algorithms for block-structured integer programs with large entries. In *Proceedings of the 2024 ACM-SIAM Symposium on Discrete Algorithms, SODA 2024, Alexandria, VA, USA, January 7 - 10, 2024*. SIAM, 2024.

11 Rodney G. Downey and Michael R. Fellows. Fixed-parameter tractability and completeness II: on completeness for W[1]. *Theor. Comput. Sci.*, 141(1&2):109–131, 1995.

12 Friedrich Eisenbrand, Christoph Hunkenschröder, Kim-Manuel Klein, Martin Koutecký, Asaf Levin, and Shmuel Onn. An algorithmic theory of integer programming. *CoRR*, abs/1904.01361, 2019.

13 Friedrich Eisenbrand and Robert Weismantel. Proximity results and faster algorithms for integer programming using the steinitz lemma. *ACM Trans. Algorithms*, 16(1):5:1–5:14, 2020. `doi:10.1145/3340322`.

14 Martin Grötschel, László Lovász, and Alexander Schrijver. *Geometric algorithms and combinatorial optimization*, volume 2. Springer Science & Business Media, 2012.

15 Raymond Hemmecke. On the positive sum property and the computation of graver test sets. *Math. Program.*, 96(2):247–269, 2003. `doi:10.1007/s10107-003-0385-7`.

16 D. S. Hochbaum and J. G. Shantikumar. Convex separable optimization is not much harder than linear optimization. *J. ACM*, 37(4):843–862, 1990.

17 A Hoffman. Integral boundary points of convex polyhedra. *Linear Inequalities and Related Systems*, pages 223–246, 1956.

18 Klaus Jansen, Kim-Manuel Klein, and Alexandra Lassota. The double exponential runtime is tight for 2-stage stochastic ILPs. *Math. Program.*, 197(2):1145–1172, 2023. `doi:10.1007/S10107-022-01837-0`.

19 Klaus Jansen and Lars Rohwedder. On integer programming, discrepancy, and convolution. *Math. Oper. Res.*, 48(3):1481–1495, 2023. `doi:10.1287/MOOR.2022.1308`.

20 Kim-Manuel Klein and Janina Reuter. Collapsing the tower - on the complexity of multistage stochastic ips. In Joseph (Seffi) Naor and Niv Buchbinder, editors, *Proceedings of the 2022 ACM-SIAM Symposium on Discrete Algorithms, SODA 2022, Virtual Conference / Alexandria, VA, USA, January 9 - 12, 2022*, pages 348–358. SIAM, 2022. `doi:10.1137/1.9781611977073.17`.

21 Dusan Knop, Martin Koutecký, Asaf Levin, Matthias Mnich, and Shmuel Onn. High-multiplicity n-fold IP via configuration LP. *Math. Program.*, 200(1):199–227, 2023. `doi:10.1007/S10107-022-01882-9`.

22 Martin Koutecký, Asaf Levin, and Shmuel Onn. A parameterized strongly polynomial algorithm for block structured integer programs. In Ioannis Chatzigiannakis, Christos Kaklamanis, Dániel Marx, and Donald Sannella, editors, *45th International Colloquium on Automata, Languages, and Programming, ICALP 2018, July 9-13, 2018, Prague, Czech Republic*, volume 107 of *LIPIcs*, pages 85:1–85:14. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2018.

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

24 Joseph Paat, Robert Weismantel, and Stefan Weltge. Distances between optimal solutions of mixed-integer programs. *Math. Program.*, 179(1):455–468, 2020. doi:10.1007/s10107-018-1323-z.

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

26 E. Steinitz. Bedingt konvergente Reihen und konvexe Systeme. *J. Reine Angew. Math.*, 146:1–52, 1916.