Collapsing the Tower—On the Complexity of Multistage Stochastic IPs

KIM-MANUEL KLEIN and JANINA REUTER, Kiel University, Germany

In this article, we study the computational complexity of solving a class of block structured integer programs (IPs), the so-called multistage stochastic IPs. A multistage stochastic IP is an IP of the form \( \min \{ e^T x \mid Ax = b, \ x \geq 0, \ x \text{ integral} \} \) where the constraint matrix \( A \) consists of small block matrices ordered on the diagonal line, and for each stage there are larger blocks with few columns connecting the blocks in a treelike fashion. Over the past few years there was enormous progress in the area of block structured IPs. For many of the known block IP classes, such as \( n \)-fold, tree-fold, and two-stage stochastic IPs, nearly matching upper and lower bounds are known concerning their computational complexity. One of the major gaps that remained, however, was the parameter dependency in the running time for an algorithm solving multistage stochastic IPs. Previous algorithms require a tower of \( t \) exponentials, where \( t \) is the number of stages. In contrast, only a double exponential lower bound was known based on the exponential time hypothesis. In this article, we show that the tower of \( t \) exponentials is actually not necessary. We show an improved running time of
\[
2^{(d\|A\|_0)O(d^{\log\log d})} \cdot \cdot r n \log (2^d)(rn)
\]
for the algorithm solving multistage stochastic IPs, where \( d \) is the sum of columns in the connecting blocks and \( r n \) is the number of rows. Hence, we obtain the first bound by an elementary function for the running time of an algorithm solving multistage stochastic IPs. In contrast to previous works, our algorithm has only a triple exponential dependency on the parameters and only doubly exponential for every constant \( t \). By this, we come very close to the known double exponential bound that holds already for two-stage stochastic IPs, i.e., multistage stochastic IPs with two stages.

The improved running time of the algorithm is based on new bounds for the proximity of multistage stochastic IPs. The idea behind the bound is based on generalization of a structural lemma originally used for two-stage stochastic IPs. While the structural lemma requires iteration to be applied to multistage stochastic IPs, our generalization directly applies to inherent combinatorial properties of multiple stages. Already a special case of our lemma yields an improved bound for the Graver complexity of multistage stochastic IPs.

CCS Concepts: • Mathematics of computing → Integer programming:

Additional Key Words and Phrases: Multistage stochastic, integer programming, parameterized complexity, stochastic programming

ACM Reference format:
Kim-Manuel Klein and Janina Reuter. 2023. Collapsing the Tower—On the Complexity of Multistage Stochastic IPs. ACM Trans. Algor. 20, 3, Article 22 (June 2023), 21 pages.
https://doi.org/10.1145/3604554

Research supported by German Research Foundation (DFG) project KL 3408/1-1.
Authors’ address: K.-M. Klein and J. Reuter, Kiel University, Christian-Albrechts-Platz 4, 24118, Kiel, Germany; emails: kmk@informatik.uni-kiel.de, janina.reuter@email.uni-kiel.de.
Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.
1549-6325/2024/06-ART22 $15.00
https://doi.org/10.1145/3604554
1 INTRODUCTION

We consider (integer) linear programs $P = (x, A, b, c)$ of the form

$$\begin{align*}
\min & \quad c^T x \\
Ax &= b \\
x &\ge 0
\end{align*}$$

for a constraint matrix $A \in \mathbb{Z}^{m_A \times n_A}$ with a specific structure, a right-hand side vector $b \in \mathbb{Z}^{m_A}$, an optimization goal $c \in \mathbb{Z}^{n_A}$, and a vector $x$ of $n_A$ variables. The constraint matrix $A$ has non-zero entries in a structure similar to Figure 1. The matrix is structured in blocks of multiple stages with the following properties. Blocks of the same stage have distinct rows and columns and for any lower stage block the set of rows is a subset of the rows of a block in the next higher stage. The subset relation on the rows induces a treelike structure as indicated by arrows.

A famous special case of multistage stochastic integer programs (IPs) are two-stage stochastic IPs, where the constraint matrix $A$ consists only of two stages, a vertical line of block matrices $A^{(i)} \in \mathbb{Z}^{t \times t'}$ and a diagonal line of block matrices $B^{(i)} \in \mathbb{Z}^{t \times t'}$, i.e.,

$$A = \begin{pmatrix}
A^{(1)} & B^{(1)} & 0 & \cdots & 0 \\
A^{(2)} & 0 & B^{(2)} & \cdots & \vdots \\
\vdots & \vdots & \ddots & \ddots & 0 \\
A^{(n)} & 0 & \cdots & 0 & B^{(n)}
\end{pmatrix}.$$

Multistage stochastic IPs appear in many real-world problems especially when problems involve uncertainty over time. This particular matrix structure models “decisions that occur at different points in time so that the problem can be viewed as having multiple stages of observations and actions” [2]. Typically either there are decisions required before all information is revealed [10, 11] or postponing decisions increases potential costs [20, 29]. A parent block models a decision and its child blocks model all scenarios that might occur in future. The quality of the decision made in a parent block depends on the occurring scenario. The areas of application include for example worker scheduling [3, 20], project planning [27, 30], routing problems [10, 11, 15], and facility location planning [29].

1.1 Related Results

Over the past few years, there has been enormous progress in the development of algorithms solving block IPs; see Table 1 for an overview. In the theoretical context, there are numerous problems where the state-of-the-art algorithm solves the problem through a block structured IP. Faster algorithms for these block structured IPs thus immediately improve the running time for algorithms for other problems. Applications include string algorithms [23], social choice games [24], scheduling [17, 18, 22], and bin packing problems [26].

Particularly useful for modelling other problems has been a block structure called $n$-fold. This block structure considers the transpose of two-stage stochastic matrices, i.e., the constraint matrix consists of a horizontal line of $n$ blocks $A^{(i)} \in \mathbb{Z}^{t \times t'}$ and a diagonal line of $n$ blocks $B^{(i)} \in \mathbb{Z}^{t \times t}$ underneath. Algorithms for $n$-fold IPs are single exponential in the block dimensions, see, e.g., Reference [5]. Though closely related, $n$-fold IPs and two-stage stochastic IPs greatly differ in their complexity. In contrast to the single exponential dependency for $n$-folds, a double exponential lower bound for the dependency on the block dimensions of two-stage stochastic IPs was recently shown under the exponential time hypothesis [16]. The lower bound is complemented by algorithms with double exponential dependency on the block dimensions [7, 9, 21].

ACM Transactions on Algorithms, Vol. 20, No. 3, Article 22. Publication date: June 2024.
Collapsing the Tower—On the Complexity of Multistage Stochastic IPs

Fig. 1. The structure of non-zero entries in a multistage stochastic matrix is denoted by filled rectangles. Rows of the blocks are connected by a tree indicated by arrows.

Table 1. In the Running Times, $\Delta$ Denotes the Maximum Absolute Entry of the Constraint Matrix $A$ and $\varphi$ Denotes the Input Length

| Overview of results for block IPs | Two-stage stochastic |
|----------------------------------|---------------------|
| $n$-fold                         |                     |
| $\Delta^{O(t(rs+st))} \cdot n^2 \varphi$ [13] | $f_1(r,s,t,\Delta) \cdot \text{poly}(n, \varphi)$ [14] |
| $(rs\Delta)^{O(r^2s+sr^2)} \cdot (nt)^2 \varphi$ [8] | $(rt\Delta)^{O(r^2s^2)} \cdot (ns)^2 \log(ns)\varphi$ [21] |
| $\Delta^{O(r^2s^2sr^2)} \cdot (nt)^3 + \varphi$ [25] | $2^{(2\Delta)^{O(r^2s+sr^2)}} \cdot n \log^3(n)\varphi$ [9] |
| $(rs\Delta)^{O(r^2s^2sr^2)} \cdot nt \log^{O(1)}(nt)\varphi^2$ [19] | $2^{(2\Delta)^{O(r^2s+sr^2)}} \cdot n^2 \log^5 n$ [9] |
| $\Delta^{O(r^2s^2sr^2)} \cdot nt \log(nt)\varphi$ [9] | $2^{(2\Delta)^{O(r^2s+sr^2)}} \cdot nt \log^2(rs)(nt)$ [7] |
| $(rs\Delta)^{O(r^2s^2sr^2)} \cdot (nt)^{1+o(1)}$ [5] |                     |

| Treefold                         | Multistage stochastic |
|----------------------------------|-----------------------|
| $f_2(t,d,r,\Delta) \cdot n^3 \varphi$ [4] | $f_5(d,\Delta,t) \cdot n^3 \varphi$ [1] |
| $(d\Delta)^{O(d^t)} \cdot (rn)^2 \log^2(rn)\varphi$ [8] | $f_4(d,r,\Delta,t) \cdot (nd)^2 \varphi \log^2(nd)$ [21] |
| $(d\Delta)^{O(d^t)} \cdot (rn)^2 \log(rn)\varphi$ [9] | $f_4(d,r,\Delta,t) \cdot n^2 \varphi$ [9] |
| $(d\Delta)^{O(d^t)} \cdot (rn)^2 \log^3(rn)$ [9] | $f_4(d,r,\Delta,t) \cdot n^3 \log^2 n$ [9] |
|                                | $f_3(d,\Delta,t) \cdot rn \log^{O(d^t)}(rn)$ [7] |

The functions $f_1$ and $f_3$ are computable and lower bounded by Ackermann’s function. Function $f_2$ is computable. The functions $f_1$ and $f_3$ involve a tower of exponents of height $t$.

The transpose of multistage stochastic matrices are treefold matrices. Multistage stochastic IPs and treefold IPs are treelike structured generalizations of two-stage stochastic IPs and $n$-fold IPs, respectively. We denote the number of stages by $t$ and the number of blocks on the lowest stage by $n$. In multistage stochastic matrices (treefold matrices), we denote the sum of column (row) dimensions for each stage by $d$ and the number of rows (columns) of the lowest stage by $r$. See also Figure 1 for the structure of multistage stochastic IPs. Recent work [8] has shown that for treefold IPs the running time dependency on the block sizes, largest matrix entry, and number of stages behaves similarly to $n$-fold IPs. It is double exponential but the second exponent only depends on the number of stages and thus the bound is single exponential for any fixed number
of stages. In contrast, prior results for two-stage and multistage stochastic IPs had a greater gap in their algorithmic complexity. The current best algorithm [7] for multistage stochastic IPs involves a tower of exponents, where the height of the tower is the number of stages. In particular, for any fixed number \( k \) of stages the dependency on the block dimensions and largest matrix entry is a tower of exponents of height \( k \).

2 OUR CONTRIBUTION

The main result of our article is to show that multistage stochastic IPs can be solved in time \( 2^{(d\Delta)^O(d^{2\omega_1})} \cdot rn \log^O(d^\Delta)(rn) \). By this, we come very close to the ETH-based double exponential lower bound of \( 2^{o(n/d)} \cdot \text{poly}(n) \) [16] that holds already for two-stage stochastic IPs. Our main ingredient to show the improved running time is a generalization of a structural lemma of Klein that is the key component in Reference [21] for the complexity bound of two-stage stochastic IPs.

**Lemma 2.1** ([21]). Consider multisets \( T_1, \ldots, T_n \subseteq \mathbb{Z}_{\geq 0}^d \) where all elements \( r \in T_i \) have bounded size \( \|r\|_\infty \leq \Delta \). Assuming that the total sum of all elements in each set is equal, i.e.,

\[
\sum_{r \in T_1} r = \cdots = \sum_{r \in T_n} r,
\]

there exist nonempty submultisets \( S_1 \subseteq T_1, \ldots, S_n \subseteq T_n \) of bounded size \( |S_i| \leq (d\Delta)^O(d\Delta^2) \) such that

\[
\sum_{s \in S_1} s = \cdots = \sum_{s \in S_n} s.
\]

The structural lemma is also applied in state-of-the-art algorithms for multistage stochastic IPs [7, 9, 21]. Simply put, the lemma describes the behavior for one stage in the multistage stochastic IP. Hence, current bounds iterate the lemma over the number of stages to obtain a bound for multistage stochastic IPs. In every iteration, the bound grows by one exponent.

Our main conceptual result is Lemma 4.1, a generalization of the structural lemma by Klein. We generalize the lemma to cope directly with arbitrarily many stages. Stating Lemma 4.1 requires a significant notational effort. However, we still want to phrase a simplified version that is notational-wise close to Lemma 2.1 to outline how it generalizes to multiple stages. For this purpose, we define partitions \( \mathcal{P}_0, \ldots, \mathcal{P}_r \) of integral numbers in the interval \([1, n]\) and any partition \( \mathcal{P}_j \) is a refinement of partition \( \mathcal{P}_{j-1} \), i.e., for every \( i \in \mathcal{P}_j \) there exists \( i' \in \mathcal{P}_{j-1} \) such that \( i \subseteq i' \). For each partition, we assign a subset of entries of \([1, \ldots, d]\), such that partition \( \mathcal{P}_i \) is assigned entries \( \{s_0 + \cdots + s_{i-1} + 1, \ldots, s_0 + \cdots + s_j\} \), with \( d = s_0 + \cdots + s_r \). See Figure 2 for an example. By \( p(i, \tau) \) we denote the projection of vector \( \tau \in \mathbb{Z}^d \) to the respective entries. In contrast to the above lemma, we do not demand equality throughout the entire sum of vectors but only within each interval in the respective vector entries of the partition.

**Lemma 2.2.** Suppose multisets \( T_1, \ldots, T_n \subseteq \mathbb{Z}_{\geq 0}^d \) are given where all elements \( r \in T_i \) have bounded size \( \|r\|_\infty \leq \Delta \) and partitions \( \mathcal{P}_0, \ldots, \mathcal{P}_r \) as described above. Assuming that the total sum of all elements in each set is equal in the assigned entries, i.e.,

\[
\sum_{r \in T_1} p(i, \tau) = \cdots = \sum_{r \in T_r} p(i, \tau) \quad \text{for every interval } I = \{\ell, \ldots, r\} \text{ in partition } \mathcal{P}_i,
\]

then there exist submultisets \( S_1 \subseteq T_1, \ldots, S_n \subseteq T_n \), which are not all empty, of bounded size \( |S_i| \leq 2^{(d\Delta)^O(d^\Delta^2)} \) such that

\[
\sum_{s \in S_1} p(i, s) = \cdots = \sum_{s \in S_n} p(i, s) \quad \text{for every interval } I = \{\ell, \ldots, r\} \text{ in partition } \mathcal{P}_i, \text{ for all } i \leq n.
\]
Collapsing the Tower—On the Complexity of Multistage Stochastic IPs

Fig. 2. An example for interval partitions with consecutive refinement, where the intervals on every horizontal level are a partition of $\{1, \ldots, 100\}$.

Using this lemma, we can show an improved bound for the size of Graver elements of a multistage stochastic matrix $A$. A Graver element is an element of $\ker\mathbb{Z}(A)$ that cannot be written as the sum of two non-zero and sign-compatible vectors in the kernel of $A$.

**Corollary 2.3.** Let $g$ be a Graver element of multistage stochastic matrix $A$. Then it is bounded by

$$
\|g\|_\infty \leq 2^{(d\Delta)^{O(d^{3r+1})}}.
$$

Using the algorithm of Eisenbrand et al. [9, Corollary 88] in combination with the improved bound for the size of Graver elements yields an algorithm for solving multistage stochastic IPs with a running time of $2^{(d\Delta)^{O(d^{3r+1})}} \cdot n^2 \varphi$, where $\varphi$ is the encoding length of the instance. The IP is of the form (1), where additionally upper and lower bounds on the variables are allowed.

We obtain furthermore a statement regarding the proximity of multistage stochastic IPs. An IP has proximity $\rho$ if for every optimal solution $x^*$ to the linear relaxation of the IP there exists an optimal integral solution $x$ such that $\|x - x^*\|_\infty \leq \rho$. Csoiovjcscek et al. [7] generalized the structural lemma of Klein such that the sums of multisets are allowed to differ slightly in the assumption. Using their generalization, they bounded the proximity of two-stage and multistage stochastic IPs. We show that a similar generalization of our Lemma 2.2 holds. By this, we derive an improved proximity bound for multistage stochastic IPs of the form (1).

**Lemma 2.4.** The proximity of multistage stochastic IPs is bounded by $2^{(d\Delta)^{O(d^{3r+1})}}$.

Our proximity bound combined with the algorithmic framework of Csoiovjcscek et al. [7] yields our main theorem.

**Theorem 2.5.** A multistage stochastic IP of the form (1) can be solved in time

$$
2^{(d\Delta)^{O(d^{3r+1})}} \cdot r n \log^{O(2^d)}(rn).
$$

3 PRELIMINARIES

For a linear program $P = (x, A, b, c)$, let $\text{Sol}_\mathbb{R}(P)$ and $\text{Sol}_\mathbb{Z}(P)$ denote the sets of fractional and integral solutions, respectively. Denote by $\text{col}(B)$ and $\text{row}(B)$ the set of column and row indices of a sub-matrix $B$ of $A$, respectively.

**Multistage stochastic matrices.** We define the shape of the constraint matrix $A$ of a multistage stochastic IP, which we will call a multistage stochastic matrix. The non-zero entries of the constraint matrix occur in a block structure with blocks $A^{(i)}, \ldots, A^{(f)}$ for $\ell \in \mathbb{Z}_{\geq 0}$, where each block uses a unique set of columns of $A$. The matrix $A$ is multistage stochastic if

- there is a block $A^{(b_i)}$ such that for every $1 \leq i \leq \ell$ we have $\text{row}(A^{(i)}) \subseteq \text{row}(A^{(b_i)})$ and
- for every two blocks $A^{(i)}, A^{(j)}$ one of the following three conditions is fulfilled: (i) $\text{row}(A^{(i)}) \subseteq \text{row}(A^{(j)})$, (ii) $\text{row}(A^{(i)}) \supseteq \text{row}(A^{(j)})$, or (iii) $\text{row}(A^{(i)}) \cap \text{row}(A^{(j)}) = \emptyset$. 

ACM Transactions on Algorithms, Vol. 20, No. 3, Article 22. Publication date: June 2024.
Multistage tree. For every multistage matrix $A$, we define a rooted tree $\mathcal{T}(A) = (V, E)$, which we call the multistage tree (of $A$). For every block $A^{(i)}$ there is a vertex $v = \text{col}(A^{(i)}) \in V$. For two vertices $u, v \in V$, where $u$ is the set of columns of $A^{(i)}$ and $v$ is the set of columns of $A^{(j)}$, row($A^{(i)}$) $\supseteq$ row($A^{(j)}$) if and only if $u$ is an ancestor of $v$. The multistage tree satisfies this condition with minimal height. An example is indicated in Figure 1.

This definition is closely related to concepts of primal treedepth of a matrix. If we consider a primal td-decomposition of the primal graph of $A$, then the multistage tree combines vertices on a path in the td-decomposition, where each vertex on the path has exactly one descendent. For more details on this topic we refer to Reference [9].

Notations. Figure 1 also visualizes parameters of the matrix. The height of the tree is denoted by $t$ and we assume that every path from the root to a leaf has the same length. We assume that vertices of the same height have the same cardinality, i.e., for every $v \in V$ of height $i$ we have $|v| = s_i$. We denote the partial sums of the number of columns by $d_i := s_0 + \cdots + s_{t-i}$. Note that then $d := d_0$ matches the primal treedepth $\text{td}_P(A)$ if the block matrices are chosen such that $d$ is minimal for the matrix $A$.

Let $n$ denote the number of leaves. We assign a number $n(v) \in \{1, \ldots, n\}$ to every leaf $v \in V$, where $n(\cdot)$ is a bijective function. For every leaf $v$ with corresponding block $A^{(i)}$ and $n(v) = i$, we define a submatrix $A_i$ of $A$. The submatrix $A_i$ consists of the entries $A_{k\ell}$, where $k \in \text{row}(A^{(i)})$ and $\ell \in \bigcup_{0 \leq h \leq t} v_h$, where $(v_0, \ldots, v_t)$ is a path from the root to leaf $v = v_i$. Let $P_i = (\bar{x}, A_i, b_i, c_i)$ denote the subprogram of the multistage stochastic program $P = (x, A, b, c)$, where $b_i$ is the projection of $b$ to the row indices row($A_i$) and $c_i$ the projection of $c$ to the column indices col($A_i$).

Throughout the article, we consider few variants of projecting vectors related to multistage structures, see also Figure 3. Define the function $\pi(i, y)$ for every $i \leq n$ and vector $y \in \mathbb{R}^{|\text{col}(A)|}$ as the projection of $y$ to the indices $\bigcup_{0 \leq j \leq t} v_j$. Note that these vectors are of dimension $d$ as $|v_i| = s_i$. Let $\pi(i, y)$ be the projection of $\pi(i, y)$ to its first $d_i$ indices, which are $\bigcup_{0 \leq k \leq t-i} v_k$. For $q \in \mathbb{R}^{d_i}$ let $\pi(q)$ be the projection to its first $d_{i+1}$ indices.

Graver bases. The conformal (partial) order $\sqsubseteq$ defined on $\mathbb{R}^d$ is satisfied by two vectors $u, v \in \mathbb{R}^n$, $u \sqsubseteq v$, if $u$ and $v$ are sign-compatible and $|u_j| \leq |v_j|$ for all components $j \in \{1, \ldots, n\}$. Two vectors $u, v \in \mathbb{R}^n$ are sign-compatible if $u_j\cdot v_j \geq 0$ for every $j \in \{1, \ldots, n\}$. The Graver basis $G(A)$ of an integral matrix $A$ consists of the inclusionwise minimal and non-zero elements of $\ker_{\mathbb{Z}}(A)$. We are interested in bounding the $\ell_p$ Graver complexity of $A$, which is defined as

$$g_p(A) := \max_{v \in G(A)} \|v\|_p$$
for any \( \rho \in [1, \infty] \). As remarked by Cslovjesc et al. [6], the classical bound for the Graver complexity [8], which depends on the number of rows of a matrix, also holds regarding the number of columns of a matrix.

**Lemma 3.1 ([6, 8]).** For every integer matrix \( A \) with \( m \) columns the Graver complexity is bounded by

\[
g_\infty(A) \leq (2m \| A \|_\infty + 1)^m.
\]

Proximity. Our proof for the proximity of multistage stochastic IPs follows the proof structure of Cslovjesc et al. [6]. Hence, we use their alternative definition and state two lemmas from their work. Proximity in this sense is a geometric measure that depends only on the polytope \( \text{Sol}_\mathbb{R}(P) \) and not on the objective function.

**Definition 1 ([6]).** Let \( P = (x, A, b, c) \) be a linear program. The proximity of \( P \), denoted \( \text{proximity}_{\infty}(P) \), is the infimum of reals \( \rho \geq 0 \) satisfying the following: For every fractional solution \( x^{\frac{\text{frac}}{\text{int}}} \in \text{Sol}_\mathbb{R}(P) \) and integral solution \( x^{\text{int}} \in \text{Sol}_\mathbb{Z}(P) \) there exists \( \tilde{x}^{\text{int}} \in \text{Sol}_\mathbb{Z}(P) \) such that

\[
\| x^{\text{int}} - x^{\frac{\text{frac}}{\text{int}}} \|_\infty \leq \rho \quad \text{and} \quad \tilde{x}^{\text{int}} - x^{\frac{\text{frac}}{\text{int}}} \subseteq x^{\text{int}} - x^{\frac{\text{frac}}{\text{int}}}.
\]

Cslovjesc et al. proved that their notion of proximity implies a bound on the usual definition of proximity. Hence, it suffices to prove bounds for \( \text{proximity}_{\infty}(A) \). Subsequently, they gave a bound for the proximity of arbitrary matrices.

**Lemma 3.2 ([6]).** Suppose \( P = (x, A, b, c) \) is a linear program. Then for every optimal fractional solution \( x^{\frac{\text{frac}}{\text{int}}} \) to \( P \) there exists an optimal integral solution \( x^{\text{int}} \) to \( P \) satisfying

\[
\| x^{\text{int}} - x^{\frac{\text{frac}}{\text{int}}} \|_\infty \leq \text{proximity}_{\infty}(P).
\]

**Lemma 3.3 ([6]).** Let \( P = (x, A, b, c) \) be a linear program where \( A \) has \( m \) columns. Then

\[
\text{proximity}_{\infty}(P) \leq (m \| A \|_\infty)^{m+1}.
\]

### 4 On the Structure of Solutions

In fact, we prove a more general lemma than Lemma 2.2, where the sums may differ slightly. Therefore, we require a notion of multisets, where the sums of elements differ slightly in the context of multistage stochastic IPs.

**Definition 2.** Consider a multistage stochastic matrix \( A \) with multistage tree \( T(A) \). Multisets \( T_1, \ldots, T_n \subset \mathbb{Z}^d \) are called \( \rho \)-valid for \( T(A) \) regarding to \( y \in \mathbb{R}^{\text{col}(A)} \) and \( \rho \in \mathbb{Z}_{>0} \) if for every \( i \leq n \) we have

\[
\left\| \sum_{\tau \in T_i} \tau - \pi(i, y) \right\|_\infty < \rho.
\]

Multisets that are 1-valid are called valid. Valid multisets regarding an integral vector are a treelike version of equivalence. A treelike equivalence of vectors in this sense means that leaves that are closely related (for example, they have the same parent) in the tree share a large portion of equal entries in the vector. Leaves that are loosely related (for example, their only common ancestor is the root) only share a fewer equal entries in the vector. A special case are two-stage stochastic IPs, where the multistage tree has height 1. In this case, valid multisets are such that the sums of vectors projected to the first \( s_0 \) indices are equal. Except for the projection, this is the same condition as in Lemma 2.1.
Valid multisets are a natural definition for multistage stochastic matrices as they yield a characterization of its integral kernel elements in the following sense. Consider any \( y \in \mathbb{Z}^{\text{col}(A)} \). If there exist multisets \( G_1 \subseteq \mathcal{G}(A_1), \ldots, G_n \subseteq \mathcal{G}(A_n) \) that are valid for \( \mathcal{T}(A) \) regarding vector \( y \), then we have that
\[
\sum_{g \in G_i} g = \pi(i, y)
\] (2)
is in the kernel of submatrix \( A_i \) for every \( i \leq n \) as it is the sum of Graver elements. Hence, \( y \) is in the kernel of \( A \). If otherwise \( y \in \ker \mathbb{Z}(A) \), then \( \pi(i, y) \) is in the kernel of submatrix \( A_i \) for every \( i \leq n \), and there exist multisets \( G_1 \subseteq \mathcal{G}(A_1), \ldots, G_n \subseteq \mathcal{G}(A_n) \) such that Equation (2) holds for every \( i \leq n \). Thus the multisets are valid for \( \mathcal{T}(A) \) regarding vector \( y \).

Observation 4.1. For a vector \( y \in \mathbb{Z}^{\text{col}(A)} \) there exist multisets \( G_1 \subseteq \mathcal{G}(A_1), \ldots, G_n \subseteq \mathcal{G}(A_n) \) that are valid for \( \mathcal{T}(A) \) regarding vector \( y \) if and only if \( y \in \ker \mathbb{Z}(A) \).

Next we will state the formal version of our main lemma. If \( \rho = 1 \) and \( y \) is integral, then this is the same statement as Lemma 2.2. The generalization to arbitrary \( \rho \in \mathbb{Z}_{>0} \) then enables us to also prove the proximity bound, as in Reference [7]. We describe the equivalence briefly. We define partitions \( \mathcal{P}_0, \ldots, \mathcal{P}_t \) of the set \( \{1, \ldots, n\} \) (the labeling of leaves in the multistage tree) as follows. For a vertex \( v \in V \), denote the set of leaves in the subtree of \( v \) by
\[
L_v = \{v' \mid v' \in V \text{ is a leaf in the subtree of } v\}.
\]
For every \( 0 \leq i \leq t \) partition \( \mathcal{P}_i \) is defined by the sets \( L_v \) of vertices \( v \in V \) of height \( i \). Clearly, this yields a partition for every height as each leaf of the tree is in the subtree of exactly one vertex of that height. Lemma 2.2 requires equality when the sums are projected on the indices of the vertex for the interval. In the following lemma, a vector \( y \) of dimension \( \text{col}(A) \) combines the indices of every vertex in the tree. The sums are compared to the corresponding components of this vector, which is equivalent to the condition in Lemma 2.2.

Lemma 4.1. Consider a multistage tree \( \mathcal{T}(A) \) and multisets \( T_1, \ldots, T_n \subseteq \mathcal{G}^d \), where each \( T_i \) contains only sign-compatible elements \( \tau \) with \( \|\tau\|_{\infty} \leq \Delta \). Assume the multisets \( T_1, \ldots, T_n \) are \( \rho \)-valid for \( \mathcal{T}(A) \) and \( \rho \in \mathbb{Z}_{>0} \) regarding a vector \( y \in \mathbb{R}^{\text{col}(A)} \). If \( \|y\|_{\infty} > \rho \cdot 2^{(d)\Delta (d+1)} \), then there exist submultisets \( S_i \subseteq T_i \), which are not all empty, and valid for \( \mathcal{T}(A) \) with respect to \( \tilde{y} \in \mathbb{Z}^{\text{col}(A)} \) with \( \|\tilde{y}\|_{\infty} \leq 2^{(d)\Delta (d+1)} \).

This is also a generalization of Theorem 9 of Cslovjecsek et al. [7]. In our notation their bound only applies to multistage trees of height 1 or in other words to two-stage stochastic matrices.

The proof of Lemma 4.1 is postponed to Section 5. Instead, this section focuses on two applications of the lemma. First, we prove our bound for the Graver complexity, and, second, we prove our bound for proximity of multistage stochastic matrices. Our proof for proximity follows the proof structure of Cslovjecsek et al. [6].

Corollary 2.3. The \( \ell_{\infty} \) Graver complexity of a multistage stochastic matrix \( A \) is bounded by
\[
g_{\infty}(A) \leq 2^{(d)\Delta (d+1)}.
\]

Proof. Let \( y \in \ker \mathbb{Z}(A) \) be a kernel element of \( A \) and assume that \( \|y\|_{\infty} > 2^{(d)\Delta (d+1)} \). By Observation 4.1, there exist multisets \( G_1 \subseteq \mathcal{G}(A_1) \) that are valid for \( \mathcal{T}(A) \) regarding \( y \), and by Lemma 3.1 the Graver complexity of the submatrices is bounded by \( g_{\infty}(A_i) \leq (2d+1)^d =: \Delta' \) for every \( 1 \leq i \leq n \).
We apply Lemma 4.1 to the multisets $G_i$, and hence there exist submultisets $S_i \subseteq G_i$ for every $i \leq n$ that are valid for $T(A)$ regarding some $\bar{y} \in \mathbb{Z}^{\text{col}(A)}$ with
\[ \|\bar{y}\|_\infty \leq 2^{(d\Delta)^{O(d^{1/\delta})}} \leq 2^{(d(2d\Delta+1)d^{1/\delta})} \leq 2^{(d\Delta)^{O(d^{1/\delta})}}. \]

Using again Observation 4.1, but now in the other direction, we get that $\bar{y} \in \ker_\mathbb{Z}(A)$. At least one submultiset $S_i$ is non-empty. The elements in each set are sign-compatible and non-zero. Hence $\pi(i, \bar{y})$ is non-zero and in particular $\bar{y}$ is. The vector $y$ is not in the Graver basis of $A$, since it is not minimal by $\bar{y} \subset y$. \hfill \Box

LEMMA 4.2. Suppose $P = (x, A, b, c)$ is a linear program and $A$ is a multistage stochastic matrix. Then
\[ \text{proximity}_{\infty}(P) \leq 2^{(d\Delta)^{O(d^{1/\delta})}}. \]

PROOF. Consider any $x^{\text{frac}} \in \text{Sol}_\mathbb{R}(P)$ and $x^{\text{int}} \in \text{Sol}_\mathbb{Z}(P)$. Let $x^{\text{int}} \in \text{Sol}_\mathbb{Z}(P)$ be an integral solution such that $x^{\text{int}} - x^{\text{frac}} \leq x^{\text{int}} - x^{\text{frac}}$ and subject to the condition that $\|x^{\text{int}} - x^{\text{frac}}\|_1$ is minimized.

If there exists a non-zero vector $u \in \ker_\mathbb{Z}(A)$ such that $u \subseteq x^{\text{frac}} - x^{\text{int}}$, then $x^{\text{int}} + u \in \text{Sol}_\mathbb{Z}(P)$ would be a solution with $(x^{\text{int}} + u) - x^{\text{frac}} \leq x^{\text{int}} - x^{\text{int}}$ and the $\ell_1$ distance from $x^{\text{frac}}$ to $x^{\text{int}} + u$ would be strictly smaller than to $x^{\text{int}}$. The existence of such a $u$ would hence contradict the choice of $x^{\text{int}}$. It is sufficient to show that in the case that $\|x^{\text{int}} - x^{\text{frac}}\|_1 > 2^{(d\Delta)^{O(d^{1/\delta})}}$ there exists a non-zero vector $u \in \ker_\mathbb{Z}(A)$ such that $u \subseteq x^{\text{frac}} - x^{\text{int}}$.

For every $i \in \{1, \ldots, n\}$ we have that $\pi(i, x^{\text{frac}}) \in \text{Sol}_\mathbb{R}(P_i)$ and $\pi(i, x^{\text{int}}) \in \text{Sol}_\mathbb{Z}(P_i)$. Every program $P_1, \ldots, P_n$ has $d$ columns. By Lemma 3.3, the proximity is bounded by $\text{proximity}_{\infty}(P_i) \leq (d\Delta)^{d+1} =: \rho$. By definition of proximity there exists an integral solution $x^{\text{int}}_i \in \text{Sol}_\mathbb{Z}(P_i)$ such that
\[ \|x^{\text{int}}_i - \pi(i, x^{\text{int}})\|_\infty \leq \rho \quad \text{and} \quad x^{\text{int}}_i - \pi(i, x^{\text{frac}}) \subseteq \pi(i, x^{\text{int}}) - \pi(i, x^{\text{frac}}). \]

We have $x^{\text{int}}_i - \pi(i, x^{\text{int}}) \in \ker_\mathbb{Z}(A_i)$, which can be decomposed into a multiset $G_i$ of Graver elements. Then $G_i$ is a multiset of sign-compatible elements of $G(A_i)$ with
\[ x^{\text{int}}_i - \pi(i, x^{\text{int}}) = \sum_{g \in G_i} g. \]

We want to apply Lemma 4.1. Therefore, we show that the multisets $G_i$ are $\rho$-valid for the multistage tree, which is the case, since
\[ \|\sum_{g \in G_i} g - \pi(i, x^{\text{frac}} - x^{\text{int}})\|_\infty = \|x^{\text{int}}_i - \pi(i, x^{\text{int}}) - \pi(i, x^{\text{frac}}) + \pi(i, x^{\text{int}})\|_\infty \leq \rho. \]

The Graver complexity of each $A_i$ is bounded by $\max_{i \leq n} g_{\infty}(A_i) \leq (2d\Delta + 1)^d$ due to Lemma 3.1. If $\|x^{\text{frac}} - x^{\text{int}}\|_\infty \geq 2^{(d\Delta)^{O(d^{1/\delta})}}$, then Lemma 4.1 can be applied and there exist non-zero submultisets $S_i \subseteq G_i$ that are valid for $T(A)$ regarding a vector $u \in \mathbb{Z}^{\text{col}(A)}$ with $\|u\|_\infty \leq 2^{(d\Delta)^{O(d^{1/\delta})}}$. By Observation 4.1, the vector $u \in \ker_\mathbb{Z}(A)$ is in the kernel of $A$.

The vector $u$ is non-zero as at least one submultiset is non-empty and every element of the multiset is non-zero and sign-compatible. Further, we have $u \subseteq x^{\text{frac}} - x^{\text{int}}$, since
\[ \pi(i, u) = \sum_{g \in S_i} g \subseteq \sum_{g \in G_i} g = x^{\text{int}}_i - \pi(i, x^{\text{frac}}) \subseteq \pi(i, x^{\text{int}}) - \pi(i, x^{\text{frac}}). \]

Due to Lemma 3.2, this bound applies to the proximity of multistage stochastic IPs in the classical sense.
5 DISCUSSION OF LEMMA 4.1

The cone, the integer cone, and the convex hull spanned by vectors \( c_1, \ldots, c_k \in \mathbb{Q}^d \) are defined by

\[
\text{cone}(c_1, \ldots, c_k) = \left\{ \sum_{i=1}^{k} \lambda_i c_i \mid \lambda \in \mathbb{R}_{\geq 0}^k \right\},
\]

\[
\text{int. cone}(c_1, \ldots, c_k) = \left\{ \sum_{i=1}^{k} \lambda_i c_i \mid \lambda \in \mathbb{Z}_{\geq 0}^k \right\}, \text{ and}
\]

\[
\text{conv}(c_1, \ldots, c_k) = \left\{ \sum_{i=1}^{k} \lambda_i c_i \mid \lambda \in \mathbb{R}_{\geq 0}^k, \|\lambda\|_1 = 1 \right\}.
\]

5.1 Proof Concept

Lemma 4.1 is a generalization of Klein’s structural lemma [21] and the stronger Klein lemma of Cslovjcesek et al. [7]. We want to give a high-level description of how our approach and in particular its connection to previous work. The proof of Klein’s structural lemma is built upon results on the intersection of integer cones. The proof by Cslovjcesek et al. refines this dependence to cones and scaling integral elements from the cone to elements in the integer cone. We rephrase two insights of their work.

Lemma 5.1 ([7]). Let \( C_1, \ldots, C_\ell \in \mathbb{Z}^{d \times d} \) be invertible integer matrices such that \( \| C_i \|_\infty \leq \Delta \) for each \( i \in \{1, \ldots, \ell\} \) and \( K := \bigcap_{i=1}^{\ell} \text{cone}(C_i) \neq \emptyset \). Then the following assertions hold:

1. The cone \( K \) can be generated by an integral matrix \( C \) with \( \| C \|_\infty \leq (d\Delta)^d \).
2. For each integer vector \( v \in K \cap \mathbb{Z}^d \) scaling by the least common multiple of determinants of the matrices yields a vector in the intersection of integer cones, i.e.,

\[
\text{lcm}_{i\leq\ell}(|\det C_i|) \cdot v \in \bigcap_{i=1}^{\ell} \text{int. cone}(C_i).
\]

They combine these insights to get elements in the intersected integer cone of \( \ell_\infty \) norm bounded by \( 2^{O(d\Delta)^d} \). This is the dominating term in the lemma as well as in the derived bounds for two-stage stochastic IPs. Bounds for multistage stochastic IPs are derived from iterating these structural lemmas over the number of stages. Eventually this leads to a tower of exponents due to \( \Delta \), the bound on the largest vector entry, being in the exponent.

It is, however, remarkable that the first step (of generating the intersected cone by integral vectors of bounded size) is much cheaper than the overall lemma. Iterating this first insight over the cones of all \( t \) stages, yields a bound of roughly \( (d\Delta)^{dO(t)} \). These integral vectors are in some sense in the intersection cone for all \( t \) stages. Using the second insight, one can get elements of the intersection integer cone of \( \ell_\infty \) norm bounded by \( 2^{(d\Delta)^{dO(t)}} \), which is only triple exponential.

Conceptually, our proof follows this idea. First, we consider only the intersection of cones over all \( t \) stages. Then, an integral vector for the intersection is scaled to an element in the intersection of integer cones over all \( t \) stages. Although the proofs of the structural results are closely related, we cannot apply this technique directly. The subset relation on the multisets is not implied by the boundedness of some elements of the intersection integer cone. Therefore, our proof requires accounting of which cone elements can be built how many times out of the given multisets. This leads us to multisets of cone elements of the intersections. An element in the intersection of high multiplicity is then scaled to an element of the intersection of integer cones over all \( t \) stages.
5.2 Accounting Integral Elements in the Intersection Cone

Let $P \subset \mathbb{Z}^d$ be the set of integral vectors bounded by $\Delta$ in infinity norm, and let $B$ denote the set of invertible $d \times d$ matrices with integral entries bounded by $\Delta$ in infinity norm. We will call elements of this set also “bases” as those will be the bases of (integer) cones. We represent a multiset $T \subseteq S$ by a multiplicity vector $\lambda \in \mathbb{R}^S$, where $\lambda_p$ denotes how often multiset $T$ contains element $p \in S$ for any set $S$. By allowing fractional multiplicity vectors, we may divide a vector $p$ into several parts, e.g., two times half of a vector $p$. When a mathematical operation on a multiplicity vector $\lambda$ requires an index $p$ that is not defined for $\lambda$, then it is treated as $\lambda_p = 0$, similarly to the fact that this vector is not included in the multiset represented by $\lambda$. For a vector $v \in \mathbb{R}^d$ and a matrix $M \in \mathbb{R}^{d \times n}$, we write $v \in M$ if $v$ is a column of $M$.

The idea presented in Subsection 5.1 requires a tool to almost partition several multisets into fractional submultisets, which each have an equal sum. Here we will prove two lemmas that provide this main tool for the proof of Lemma 4.1. The following lemma, Lemma 5.2, shows the existence of an integral element in the intersection cone that can be represented as a fractional submultiset of every multiset $i \leq n$. Such an element $\hat{q}$ is represented by only vectors of one basis $B$ and by a fractional multiplicity vector $\mu$ for that basis, in particular $B\mu = \hat{q}$. Every vector $p \in B$ is used $\mu_p$ times in the representation. As $\mu$ is uniquely defined by $\mu = B^{-1}\hat{q}$, we will treat $B^{-1}\hat{q}$ as a multiplicity vector for the column vectors of $B$. The proof works similar to the stronger Klein bound in Reference [6].

**Lemma 5.2.** Consider multiplicity vectors $\lambda^{(1)}, \ldots, \lambda^{(m)} \in \mathbb{Q}_{\geq 0}^P$, a vector $q \in \mathbb{R}^d$, and $\rho \in \mathbb{Z}_{\geq 0}$ such that for every $i \leq m$ we have

$$\left\| \sum_{p \in P} \lambda^{(i)}_p p - q \right\|_{\infty} < \rho.$$ 

If $\|q\|_{\infty} > \rho \cdot (d\Delta)^O(d^2)$, then there exist bases $B^{(1)}, \ldots, B^{(m)} \in B$ and $\hat{q} \in \mathbb{Z}^d$ with $\hat{q} \neq 0$ such that

(i) $0 \leq \mu^{(i)} \leq \lambda^{(i)}$ for $\mu^{(i)} := (B^{(i)})^{-1}\hat{q}$,
(ii) $\|\hat{q}\|_{\infty} \leq (d\Delta)^{d^2}$.

**Proof.** For every $i \leq m$ let $r^{(i)} \in \mathbb{R}^P_{\geq 0}$ be such that

$$\sum_{p \in P} \left( \lambda^{(i)}_p + r^{(i)}_p \right) p = q$$

and w.l.o.g. $\|r^{(i)}\|_{\infty} \leq \rho$ using only the unit vectors. Define the sum $z^{(i)} := \lambda^{(i)} + r^{(i)}$ for every $i \leq m$. Every $z^{(i)}$ belongs to the polyhedron $Q = \{ \mu \in \mathbb{R}^P_{\geq 0} \mid \sum_{p \in P} \mu_p p = q \}$. By the Minkowski-Weyl theorem [28], the polyhedron can be written as

$$Q = \text{conv}(u^{(1)}, \ldots, u^{(\ell)}) + \text{cone}(v^{(1)}, \ldots, v^{(k)})$$

for some $u^{(1)}, \ldots, u^{(\ell)}, v^{(1)}, \ldots, v^{(k)} \in \mathbb{R}^P_{\geq 0}$, where $u^{(j)} \geq 0$ and $\sum_{p \in P} u^{(j)}_p p = q$ and $v^{(j)} \geq 0$ and $\sum_{p \in P} v^{(j)}_p p = 0$.

Every $u^{(j)}$ is a vertex solution to the linear program of $Q$ and has at most $d$ non-zero entries, and hence there exists an invertible matrix $C^{(j)}$ such that $C^{(j)} u^{(j)} = q$. By Carathéodory’s theorem, every $z^{(i)}$ can be written as

$$z^{(i)} = \sum_{j=1}^\ell Y_j u^{(j)} + \sum_{k=1}^p \mu_k v^{(k)},$$

where $\sum_{j=1}^\ell Y_j = 1$ and $Y_j, \mu_k \in \mathbb{R}_{\geq 0}$ and $Y_j$ is nonzero for at most $d + 1$ indices $j$. Thus, by the pigeonhole principle there exists an index $j(i)$ with $Y_{j(i)} \geq \frac{1}{d+1}$. Since all scalars and vectors are
non-negative, we have
\[ 0 \leq \frac{1}{d+1} u^{(i)}(i) \leq z^{(i)}. \] (3)

Consider the intersection \( \bigcap_{i=1}^{\ell} \text{cone}(C^{(i)}) \), which is a cone with some minimal generating set \( C \subseteq \mathbb{Z}^d \). A consequence of the Farkas-Minkowski-Weyl theorem, see, e.g., Reference [28, Corollary 7.1a], is that the generating elements can be bounded by \( \|c\|_\infty \leq (d\Delta)^{d^2} \) for every \( c \in C \) as described in Reference [6].

For every \( i \leq \ell \) we have that \( q \in \text{cone}(C^{(i)}) \). Hence, the vector is also in the intersection \( q \in \text{cone}(C) \) and there exist by Carathéodory’s theorem \( d \) vectors \( c^{(1)}, \ldots, c^{(d)} \in C \) with
\[ q = \sum_{k=1}^{d} \alpha_k c^{(k)} \quad \text{for some } \alpha \in \mathbb{R}_{\geq 0}. \]

For the purpose of this proof we can assume that
\[ \|q\|_\infty > \rho \cdot (d\Delta)^O(d^2) = (d+1)d \cdot 2\rho \max_i \|c^{(i)}\|_\infty \max_i |\det(C^{(i)})|, \]
where \( \max_i |\det(C^{(i)})| \leq (d\Delta)^d \) by the Hadamard bound. By the pigeonhole principle, there exists \( \alpha_k > 2\rho(d+1) \cdot \max_i |\det(C^{(i)})| \), and without loss of generality assume \( k = 1 \). For each \( j \in \{1, \ldots, \ell\} \) there exist \( y^{(j)}, \tilde{y}^{(j)} \in \mathbb{R}_{\geq 0}^d \) with \( C^{(j)} y^{(j)} = c^{(1)} \) and \( C^{(j)} \tilde{y}^{(j)} = -\frac{q}{(d+1)} - 2\rho \max_i |\det(C^{(i)})| \cdot c^{(1)} \). We can write
\[ C^{(j)} u^{(j)}/(d+1) = q/(d+1) = C^{(j)} \left( 2\rho \max_i |\det(C^{(i)})| \cdot y^{(j)} + \tilde{y}^{(j)} \right). \]

We have that \( 2\rho \max_i |\det(C^{(i)})| \cdot y^{(j)} \leq u^{(j)}/(d+1) \), since \( C^{(j)} \) is invertible and \( \tilde{y}^{(j)} \) is non-negative.

We set \( \hat{q} := c^{(1)} \) and \( \hat{y}^{(i)} := C^{(j)} \). The size of the vector is bounded by \( \|\hat{q}\|_\infty = \|c^{(1)}\|_\infty \leq (d\Delta)^{d^2} \), which is property (i). For the submultiset relation (i) recall that
\[ 0 \leq 2\rho \max_i |\det(C^{(i)})| \cdot y^{(j)(i)} \leq \frac{1}{d+1} u^{(j)(i)} \leq \frac{1}{d+1} z^{(i)} = \lambda^{(i)} + r^{(i)}. \]

Thus we have that
\[ y^{(j)(i)} \leq \frac{1}{2\rho \max_i |\det(C^{(i)})|} \lambda^{(i)} + \frac{1}{2\rho \max_i |\det(C^{(i)})|} r^{(i)}. \]

By Cramer’s rule the denominator of \( y^{(j)(i)} \) is at most \( |\det(C^{(j)})| \) for all \( k \leq d \). As \( \|r^{(i)}\|_\infty \leq \rho \), we have that
\[ y^{(j)(i)} \leq \frac{1}{2\rho \max_i |\det(C^{(i)})|} \lambda^{(i)} \leq \lambda^{(i)}. \]

In conclusion, this shows that \( (B^{(i)})^{-1} \hat{q} = (C^{(j)})^{-1} c^{(1)} = y^{(j)(i)} \leq \lambda^{(i)} \).\( \square \)

To compare the rough idea from Subsection 5.1 to Lemmas 5.2 and Lemma 5.3, Lemma 5.2 in some sense finds an integral element \( \hat{q} \) from the intersection cone. In other words, Lemma 5.2 describes the terms on which we can build an integral cone element from the given multisets. Lemma 5.3 then accounts of which cone elements can be build how many times out of the given multisets. We denote the bookkeeping of which integral element in the intersection cone can be constructed how many times from the multisets again as a multiset.

This multiset of integral elements in the intersection is obtained by iterating Lemma 5.2. Every iteration yields an element, \( \hat{q} \) in the above lemma, that is added to the new multiset. Each element represents a submultiset for every \( i \) as it is a fractional combination of elements from the original
multisets. Property (i) ensures the submultiset relation. Let $P' \subset \mathbb{Z}^d$ be the set of integral vectors bounded by $(d\Delta)^d$ in infinity norm.

**Lemma 5.3.** Consider multiplicity vectors $\lambda^{(1)}, \ldots, \lambda^{(m)} \in \mathbb{Q}^P_{\geq 0}$ and a vector $q \in \mathbb{R}^d$ such that for every $i \leq m$ we have

$$\left\| \sum_{p \in P} \lambda_p^{(i)} p - q \right\|_{\infty} < \rho.$$ 

There exist multiplicity vectors $\lambda[B, i] \in \mathbb{Z}^P_{\geq 0}$ for every $B \in \mathcal{B}$ and $i \leq m$ such that

1. $\lambda := \sum_{B \in \mathcal{B}} \lambda[B, 1] = \cdots = \sum_{B \in \mathcal{B}} \lambda[B, m]$,
2. $\sum_{\rho \in P'} \sum_{B \in \mathcal{B}} \lambda[B, i]_p (B^{-1}p) \leq \lambda^{(i)}$ for every $i \leq m$,
3. if $\lambda[B, i]_p > 0$, then $(B^{-1}p) \geq 0$ for every $B \in \mathcal{B}$, $p \in P$, and $i \leq m$,
4. $\left\| \sum_{p \in P'} \lambda_p^{(i)} p - q \right\|_{\infty} \leq \rho \cdot (d\Delta)^{O(d^2)}$.

**Proof.** We want to iterate Lemma 5.2. To start the iteration let $\lambda^{(i)}[0] = \lambda^{(i)}$ for $i \leq m$ and $\hat{q}[0] := q$. If in iteration $j$ we have that $\|\hat{q}[j]\|_{\infty} > \rho \cdot (d\Delta)^{O(d^2)}$, then we apply Lemma 5.2 to the multiplicity vectors $\lambda^{(i)}[j]$, and there exists bases $B^{(1)}[j], \ldots, B^{(m)}[j] \in \mathcal{B}$ and $\hat{q}[j] \in P'$ as stated in Lemma 5.2. For the next iteration, we define $\hat{q}[j + 1] := \hat{q}[j] - \hat{q}[j]$ and $\lambda^{(i)}[j + 1] := \lambda^{(i)}[j] - ((B^{(i)}[j])^{-1}\hat{q}[j])$, where the latter is non-negative by Lemma 5.2(i). We have that

$$\left\| \sum_{p \in P'} \lambda_p^{(i)}[j + 1] p - \hat{q}[j + 1] \right\|_{\infty} = \left\| \sum_{p \in P'} \lambda_p^{(i)}[j] p - \hat{q}[j] \right\|_{\infty} \leq \rho.$$ 

Let $v \in \mathbb{Z}_{\geq 0}$ be the number of iterations until the condition $\|\hat{q}[j]\|_{\infty} > \rho \cdot (d\Delta)^{O(d^2)}$ does not hold. In particular, the $\ell_\infty$ norm of $\hat{q}[v]$ is bounded by $\|\hat{q}[v]\|_{\infty} \leq \rho \cdot (d\Delta)^{O(d^2)}$.

For every $B \in \mathcal{B}$, $i \leq m$, and $\hat{q} \in P'$ denote by

$$\lambda[B, i]_{\hat{q}} = |\{j \leq v \mid \hat{q}[j] = \hat{q} \text{ and } B^{(i)}[j] = B\}|$$

how often basis $B$ was used for multiset $i$ and vector $\hat{q}$. By Lemma 5.2(i) we have $(B^{-1}p) \geq 0$ if $\lambda[B, i]_p > 0$ for every $B \in \mathcal{B}$, $p \in P$, and $i \leq m$. This is property (iii). By definition, we have for every $\hat{q} \in P'$

$$|\{j < v \mid \hat{q}[j] = \hat{q}\}| = \sum_{B \in \mathcal{B}} \lambda[B, 1]_{\hat{q}} = \cdots = \sum_{B \in \mathcal{B}} \lambda[B, m]_{\hat{q}}.$$

For every $i \leq m$, we have property (ii) by

$$\lambda^{(i)} = \sum_{j < v} (B^{(i)}[j])^{-1}\hat{q}[j] + \hat{\lambda}^{(i)}[v] \geq \sum_{j < v} (B^{(i)}[j])^{-1}\hat{q}[j] = \sum_{\hat{q} \in P'} \sum_{B \in \mathcal{B}} \lambda[B, i]_{\hat{q}} \cdot (B^{(i)})^{-1}\hat{q}.$$

For property (iv), the sum for the constructed multisets and $q$ differ only by the remaining vector $\hat{q}[v]$, since $\sum_{p \in P'} \lambda_p P = \sum_{j < v} \hat{q}[j]$ and

$$\left\| \sum_{p \in P'} \lambda_p p - \hat{q} \right\|_{\infty} = \left\| \sum_{p \in P'} \lambda_p p - \left( \sum_{j < v} \hat{q}[j] + \hat{q}[v] \right) \right\|_{\infty} \leq \|q[v]\|_{\infty} \leq \rho \cdot (d\Delta)^{O(d^2)}.$$
5.3 Proof of Lemma 4.1

Lemma 5.3 almost partitions the given multisets into submultisets representing integral elements of the intersection cone. In terms of our proof concept, this provides us with one round of intersecting cones. In the proof of Lemma 4.1, we will use Lemma 5.3 to construct multisets for every vertex in the multistage tree. For vertices of height $0 \leq j \leq t$, the elements will be bounded by $\Delta_j := (d\Delta)^{d^j}$ in the $\ell_\infty$ norm. We consider the underlying set

$$
P_j := \{ p \in \mathbb{Z}^{d_j} \mid \|p\|_\infty \leq \Delta_j \}
$$

and a variant, where elements are of lower dimension,

$$
\hat{P}_j := \pi(P_j) = \{ p \in \mathbb{Z}^{d_{j+1}} \mid \|p\|_\infty \leq \Delta_j \}.
$$

As before, elements in newly constructed multisets are represented by a certain subset of $\hat{P}_j$ that happens to be a basis. We thus also define the set

$$
\mathcal{B}_j := \{ B \in \mathbb{Z}^{d_{j+1} \times d_{j+1}} \mid B \text{ is a basis with } \|B\|_\infty \leq \Delta_j \}.
$$

Let $K_1$ be the constant in the $O$-notation of Lemma 5.3(iv). We will further show that the respective sums for every child differ from the initial vector $y$ by at most $\rho_j := \rho \cdot (d\Delta)^{K_1 d^j}$ for every height $0 \leq j \leq t$. Recall that $\pi^j(i, y)$ is the projection of $\pi(i, y)$ to its first $d_j$ indices and $\pi(i, y)$ is the projection to indices $\{ j \leq t \}$ if $(v_0, \ldots, v_t)$ is the path from the root $v_0$ to a leaf $v_t$ with $v_i = i$ for all $i \leq n$ and $y \in \mathbb{R}^{\text{col}(A)}$. If $q \in \mathbb{R}^{d_j}$, then $\pi(q)$ is its projection on the first $d_{j+1}$ indices.

**Lemma 4.1.** Consider a multistage tree $T(A)$ and multisets $T_1, \ldots, T_n \subset \mathbb{Z}^d$, where each $T_k$ contains only sign-compatible elements $\tau$ with $\|\tau\|_\infty \leq \Delta$. Assume the multisets $T_1, \ldots, T_n$ are $q$-valid for $T(A)$ regarding a vector $y \in \mathbb{R}^{\text{col}(A)}$.

If $\|y\|_\infty > \rho \cdot 2^{(d\Delta)^{O(d^{j+1})}}$, then there exist submultisets $S_i \subset T_i$, which are not all empty, and valid for $T(A)$ with respect to $\hat{y} \in \mathbb{Z}^{\text{col}(A)}$ with $\|\hat{y}\|_\infty \leq 2^{(d\Delta)^{O(d^{j+1})}}$.

**Proof Sketch for Lemma 4.1.** The proof consists of two phases. The first phase considers the multistage tree in a bottom up fashion and constructs multisets for every vertex of the tree using the partitioning lemma, Lemma 5.3. For a vertex $v \in V$ of height $j \leq t$, vectors in the multiset have dimension $d_j$. In principle, this is the intersection of cones over the $t$ stages and dimensions but with bookkeeping. Analyzing the relation of $\|y\|_\infty$ and $\|\pi^j(i, y)\|_\infty$, we arrive at either the case that the multiset constructed for the root contains an element of sufficient high multiplicity or, in the other case, there exists a vertex where the multiset contains the vector $0$ sufficiently often. In both cases, the second phase uses the high multiplicity element to reconstruct submultisets for every vertex in the subtree in a top down fashion. In comparison to the proof concept, this step is much like scaling an integral vector from the intersection cone to a vector in the intersection integer cone. The reconstruction maintains the important invariant that the sum of elements in the submultisets remains the same for an index once it is defined. One might think of the reconstruction as starting from the indices for $\pi^j(i, \hat{y})$, and in each step the vector $\hat{y}$ is extended by the indices $i \in v'$ for child vertices $v'$. By this invariant, the constructed submultisets are valid for the subtree for every step. At last, the submultisets that are constructed for the leaves of $T(A)$ are valid for the multistage tree.

We want to preview some technical details. For every vertex $v \in V$, a multiset $\lambda^v$ is considered. Another multiset $\hat{\lambda}^v$ considers the projection $\pi(.)$ of the multiset $\lambda^v$. In other words, $\hat{\lambda}^v$ is the multiset obtained when every element $p$ of $\lambda^v$ is projected to $\pi(p)$. In the second phase, the lemma constructs valid submultisets from an element of $\lambda^v$ with high multiplicity. The reconstruction starting at a vertex $v$ of height $0 \leq j \leq t$ requires an element $p$ of multiplicity $\lambda_p^v \geq \alpha_j \cdot \beta_j$ for
some $\alpha_j, \beta_j \in \mathbb{Z}_{\geq 0}$ that are defined in the proof below. The reconstruction then extends the vector $\alpha_j \cdot p$, where scaling the vector $p$ with $\alpha_j$ will be used to scale from a fractional number of vectors to an integral number of vectors. In particular, Lemma 5.3 is used to fractionally partition multisets to obtain a multiset for the parent, but scaling $p$ with $\alpha_j$ ensures that we find an extension of $\alpha_j \cdot p$ that is an integral combination of elements in $T_i$ for every $i \leq n$. However $\beta_j$ leaves room between the multiplicities used for the reconstruction and those available in the multiset to use pigeonhole principle when needed.

Proof of Lemma 4.1. Constructing multisets for the tree. For all vertices $v \in V$, we construct multisets as follows. If $v$ is a leaf with $n(v) = i$, then let $\lambda^v \in \mathbb{Z}_{\geq 0}^{P_0}$ be the multiplicity vector representation of multiset $T_i$. This representation is possible, since $T_i \subset P_0$ as every $\tau \in T_i$ has dimension $d_0$ and is bounded by $\|\tau\|_\infty \leq \Delta \leq \Delta_0$. Let $\hat{\lambda}^v \in \mathbb{Z}_{\geq 0}^{P_0}$ be defined for every $\hat{p} \in \hat{P}_0$ by

$$\hat{\lambda}^v_p := \sum_{\substack{\rho \in P_0 \\text{s.t.} \\pi(\rho) = \hat{p}}} \lambda^v.$$

(5)

Note that $\| \sum_{\rho \in P_0} \hat{\lambda}^v_p - \pi^i(i, y) \|_\infty \leq \| \sum_{\tau \in T_i} \tau - \pi(i, y) \|_\infty \leq \rho = \rho_0$, since the multisets $T_1, \ldots, T_n$ are $\rho$-valid for $T(A)$.

Consider an inner vertex $v \in V$ of height $j \leq t$, where for all children $v_1, \ldots, v_m$ the multiplicity vectors $\lambda^{v_k} \in \mathbb{Z}^{P_{j-1}}$ and $\hat{\lambda}^{v_k} \in \mathbb{Z}^{\hat{P}_{j-1}}$ were already defined and, for all $1 \leq i \leq n$,

$$\left\| \sum_{\rho \in P_{j-1}} \hat{\lambda}^{v_k}_p \rho - \pi^j(i, y) \right\|_\infty \leq \rho_{j-1}$$

for all $k \leq m$. We apply Lemma 5.3 on the multiplicity vectors $\hat{\lambda}^{v_1}, \ldots, \hat{\lambda}^{v_m}$. The multiplicity vector $\hat{\lambda}^{v_k}$ is indexed by $\hat{P}^{j-1}$, and thus the $\ell_{\infty}$ norm of elements from $\hat{\lambda}^{v_k}$ is bounded by $\Delta_{j-1}$. The $\ell_{\infty}$ norm of elements in the multiplicity vectors indexed by $P'$ in Lemma 5.3 is hence bounded by

$$(d_j \Delta_{j-1})^{d_{j}} = (d_j (d\Delta)^{d_{j-1}})^{d_{j}} \leq (d\Delta)^{d_{j}} = \Delta_j.$$

The multiplicity vectors indexed by $P'$ in Lemma 5.3 can therefore be indexed by $P_j$ in this context. By Lemma 5.3, there exist multiplicity vectors $\lambda[B, v_k] \in \mathbb{Z}^{P_{j-1}}$ for every $B \in B^{j-1}$ and $k \leq m$ such that

(i) $\lambda^v := \sum_{B \in B^{j-1}} \lambda[B, v_1] = \cdots = \sum_{B \in B^{j-1}} \lambda[B, v_m],$

(ii) $\sum_{p \in P_j} \sum_{B \in B^{j-1}} \lambda[B, v_k] (B^{-1} p) \leq \lambda^{v_k}$ for every $k \leq m$,

(iii) if $\lambda[B, v_k]_p > 0$, then $(B^{-1} p) \geq 0$ for every $B \in B^{j-1}$, $p \in P_j$, and $k \leq m$,

(iv) $\| \sum_{p \in P_j} \hat{\lambda}^v_p \rho - \pi^j(i, y) \|_\infty \leq \rho_{j-1} \cdot (d\Delta_{j-1})^{O(d_{j})}$.

We define the multiplicity vector $\hat{\lambda}^v \in \mathbb{Z}^{\hat{P}_j}$ similarly to the leaves in Equation (5). Observe that by Equation (iv) for all $1 \leq i \leq n$ we have that

$$\left\| \sum_{p \in P_j} \hat{\lambda}^v_p \rho - \pi^j(i, y) \right\|_\infty \leq \rho_{j-1} \cdot (d_j \Delta_{j-1})^{K_{j}d_{j}} \leq \rho \cdot (d\Delta_{j})^{K_{j}d_{j}} = \rho_j.$$  

The relation of $\|\gamma\|_\infty$ and $\|\pi^j(i, y)\|_\infty$. Define $\nu := \text{lcm}(1, \ldots, (d\Delta_{t-1})^{d_{j}}$, $\alpha_i := \nu^i$, and $\beta_i := \Delta_{t}^{2i}d_{i}$. We will focus on the case that for every $v \in V$ of height $j$, we have $\hat{\lambda}^v_0 \leq D_j$ for $D_j := \alpha_i \cdot \beta_i$. The other case, that is, $\hat{\lambda}^v_0 > D_j$ for some vertex $v \in V$, is discussed at the end of the proof.
Let \( v \in V \) be a vertex of height \( j \) and consider a child \( v' \in V \). Due to \( \| \sum_{p \in P_{j-1}} \hat{\lambda}'_p p - \pi^j(i, y) \|_\infty \leq \rho_{j-1} \) and \( \| \sum_{p \in P_j} \lambda^j_p p - \pi^j(i, y) \|_\infty \leq \rho_j \) we have that
\[
\left\| \sum_{p \in P_{j-1}} \hat{\lambda}'_p p - \sum_{p \in P_j} \lambda^j_p p \right\|_\infty \leq \rho_j + \rho_{j-1}.
\]
By reverse triangle inequality we get that
\[
\left\| \sum_{p \in P_{j-1}} \hat{\lambda}'_p p \right\|_\infty \leq \sum_{p \in P_{j-1}} \lambda^j_p p + \rho_j + \rho_{j-1} \leq \Delta_j \| \hat{\lambda}' \|_1 + \rho_j + \rho_{j-1}. \tag{6}
\]
To compare the \( \ell_1 \) norms of \( \hat{\lambda}' \) and \( \hat{\lambda}' \), we consider the left part of the above inequality. The multisets \( T_i \) contain only sign-compatible vectors, which implies
\[
\left\| \sum_{p \in P_{j-1}} \hat{\lambda}' p \right\|_\infty = \sum_{p \in P_{j-1}} \hat{\lambda}' p \geq \frac{\sum_{p \in P_{j-1} \setminus \{0\}} \hat{\lambda}' p \hat{\gamma} p |_{D_j} \geq \frac{\| \hat{\lambda}' \|_1 - D_{j-1}}{d_j}. \tag{7}
\]
As a combination of Equations (6) and (7) and using \( d_{j-1} \geq d_j \), it holds that
\[
\frac{\| \hat{\lambda}' \|_1 - D_{j-1}}{d_{j-1}} \leq \Delta_j \| \hat{\lambda}' \|_1 + \rho_j + \rho_{j-1} \leq \Delta_j \left( \sum_{p \in P_{j-1} \setminus \{0\}} \hat{\lambda}' p + \Delta_j \right) + \rho_j + \rho_{j-1}. \tag{8}
\]
To compare the \( \ell_\infty \) norms of \( \pi^j(i, y) \) and \( \pi^{j-1}(i, y) \), they are compared to the \( \ell_1 \) norms of \( \hat{\lambda}' \) and \( \hat{\lambda}' \), respectively. Note that the \( \ell_1 \) norms of \( \hat{\lambda}' \) and \( \hat{\lambda}' \) are equal by definition. Using the distance to \( y \) we get
\[
\| \hat{\lambda}' \|_1 \geq \frac{\| \pi^{j-1}(i, y) \|_\infty - \rho_{j-1}}{\Delta_{j-1}} \text{ and } \sum_{p \in P_{j-1} \setminus \{0\}} \hat{\lambda}' p \leq d_{j-1}(\| \pi^j(i, y) \|_\infty + \rho_j). \tag{9}
\]
Combining the above with Equation (8) yields
\[
\| \pi^{j-1}(i, y) \|_\infty \leq \Delta_j^2 d_{j-1}^{2t}(\| \pi^j(i, y) \|_\infty + 2\rho_j + 2\rho_{j-1} + D_j + D_{j-1})
\leq \Delta_j^2 d_t^2(\| \pi^j(i, y) \|_\infty + 4\rho_j + 2D_t).
\]
For every \( i \leq n \) by induction it holds that
\[
\| y \|_\infty \leq \Delta_t^2 d_t^2(\| \pi^t(i, y) \|_\infty + 4t\rho_j + 2tD_t).
\]
**Reconstructing submultisets.** We remain in the case that \( \lambda^j_0 \leq D_j \) for every vertex. Let \( r \) be the root of the multistage tree. There exists \( p \in P_t \setminus \{0\} \) with
\[
\lambda^r p \geq \frac{1}{\Delta_t^2} \| \hat{\lambda}^r \|_1 \geq \frac{1}{\Delta_t^2} \sum_{p \in P_t \setminus \{0\}} \lambda^r p \geq \frac{1}{\Delta_t^{d+1}}(\| \pi^t(i, y) \|_\infty - \rho_t).
\]
For the purpose of this proof, we can assume that
\[
\| y \|_\infty \geq \Delta_t^{2t} d_t^{2t} (\Delta_t^{d+1} \alpha_t \beta_t + \rho_t + 4t\rho_j + 2tD_t) = \rho : 2^{(dA \bar{\alpha}(d^t))}
\]
holds and thus \( \| \pi^t(i, y) \|_\infty \geq \Delta_t^{d+1} \alpha_t \beta_t + \rho_t \) for all \( i \leq n \) and \( \lambda^r p \geq \alpha_t \beta_t \). We reconstruct submultisets from the root to every vertex and finally for the leaves. We start with the multiplicity vector \( y^r := e_p \) for the root \( r \), where \( e_p \) is the unit vector with value 1 at index \( p \).
Claim: Consider a vertex \( v \in V \) of height \( 1 \leq j \leq t \). Then for every \( v' \in \mathbb{Z}_{\geq 0}^{P_j} \) with \( \alpha_j \beta_j \cdot v' \leq \lambda v \) there exists \( y^{v'} \in \mathbb{Z}_{\geq 0}^{P_{j-1}} \) for every child \( v' \in V \) such that

(i) \( \alpha_{j-1} \beta_j \cdot y^{v'} \leq \lambda v' \),
(ii) \( \alpha_{j-1} \| y^{v'} \|_1 \leq d (d \Delta_t)^d \alpha_j \| y^{v} \|_1 \), and
(iii) \( \pi (\alpha_{j-1} \sum_{p \in P_j} y_p') \) = \( \alpha_j \sum_{p \in P_j} y_{v'} \).

Proof of the Claim: Consider \( v' \in \mathbb{Z}_{\geq 0}^{P_j} \) with \( \alpha_j \beta_j y_p' \leq \lambda v' \) for every \( p \in P_j \). To prove the claim, first the basis representation of the multiplicity vector \( \lambda v' \) for every child \( v' \) is considered and a basis of sufficient multiplicity \( \lambda [B, v'] \) is found for each vector in the multiplicity vector. Then the vectors are extended to dimension \( d_{j-1} \) using a vector with sufficient multiplicity in \( \lambda v' \). Finally, the properties of the claim are verified.

We want to use the representation of the elements in \( \lambda v \) to obtain elements in \( \lambda v' \) for any child \( v' \). Each representation is defined by the used basis. Hence, we start with the basis representation. By definition, see Equation (i), we have

\[
\lambda v = \sum_{B \in \mathcal{B}^{j-1}} \lambda [B, v']
\]

for every child \( v' \in V \) of \( v \). There are at most \( \Delta_t^{d_j} \) bases \( B \) with a weight \( \lambda [B, v'] \neq 0 \), because every multiset \( T_i \) contains sign-compatible elements and this is preserved for \( \lambda v' \). Hence, for every \( p \in P_j \) there exists a basis \( B^p \in \mathcal{B}^{j-1} \) with \( \lambda v' \leq \lambda [B^p, v'] \) by pigeonhole principle. Hence we get

\[
\frac{1}{\Delta_t^{d_j}} \alpha_j \beta_j y_p' \leq \frac{1}{\Delta_t^{d_j}} \lambda v' \leq \lambda [B^p, v']
\]

(10)

After we selected a basis for every vector, we want to use that the basis and the vector represent a fractional submultiset. To combine the chosen representations in a new multiplicity vector, we define a vector \( \hat{y}^{v'} \in \mathbb{Z}_{\geq 0}^{P_{j-1}} \) by

\[
\hat{y}^{v'} := v \cdot \sum_{p \in P_j} y_p' \cdot (B^p)^{-1} \cdot p.
\]

The definition is simply the sum of how many times we require that element of the child multiset to represent our current multiset. Note that \( \hat{y}^{v'} \in \mathbb{Z}_{\geq 0} \) for every \( p' \in \mathcal{P}_{j-1} \), since all vectors and matrices are integral, by Cramer’s rule \( (B^p)^{-1} \), has denominators at most \( | \text{det}(B^p) | \leq (d \Delta_t)^d \) that divides \( v \) and \( (B^p)^{-1} p \geq 0 \) by Lemma 5.3(iii).

Again, we verify that this representation is a submultiset of the multiset of each child. By Lemma 5.3(ii), the following inequality is given:

\[
\sum_{p \in P_j} \sum_{B \in \mathcal{B}^{j-1}} \text{s.t. } p' \in B \lambda [B, v'] \cdot ((B^p)^{-1} \cdot p) \leq \hat{y}^{v'}.
\]

(11)

Using Equation (11), the bound in Equation (10) can be extended to \( \hat{y}^{v'} \) and \( \hat{\lambda}^{v'} \) as follows:

\[
\frac{1}{\Delta_t^{d_j}} \alpha_j \beta_j \cdot y_p' = \frac{1}{\Delta_t^{d_j}} \alpha_j \beta_j \cdot v \cdot \sum_{p \in P_j} y_p' \cdot (B^p)^{-1} \cdot p
\]

\[
\leq \sum_{p \in P_j} \lambda [B^p, v'] \cdot ((B^p)^{-1} \cdot p)
\]

(10)

\[\lambda [B^p, v'] \cdot ((B^p)^{-1} \cdot p]

ACM Transactions on Algorithms, Vol. 20, No. 3, Article 22. Publication date: June 2024.
\[ \leq \sum_{p \in P} \sum_{B \in \mathcal{B}^j \text{ s.t. } p' \in B} \lambda[B, v']_p ((B^p)^{-1} p)_{p'} \]

\[ \leq \hat{\lambda}' \hat{v}'. \quad (11) \]

Next, each vector is extended to dimension \( d_{j-1} \) to reverse the projection from \( \lambda' \) to \( \hat{\lambda}' \). For every \( p \in \hat{P}^{j-1} \), there are at most \( \Delta^j_{j-1} \) sign-compatible vectors \( p'' \in P^{j-1} \) that are projected to \( p' \), i.e., \( \pi(p'') = p' \), and by definition

\[ \hat{\lambda}'_{p'} = \sum_{p'' \in P^{j-1}} \lambda_{p''} \quad \text{s.t. } \pi(p'') = p'. \]

Hence, for every \( p' \in \hat{P}^{j-1} \) there exists \( p'' \in P^{j-1} \), where \( \frac{1}{\Delta_t} \hat{\lambda}'_{p'} \leq \lambda'_{p''} \) and

\[ \frac{1}{\Delta_t} \hat{\lambda}'_{p'} \leq \lambda'_{p''}. \quad (13) \]

Let \( f : \hat{P}^{j-1} \to P^{j-1} \) map any \( p' \in \hat{P}^{j-1} \) to some \( p'' \in P^{j-1} \) such that \( \pi(p'') = p' \) and \( p'' \) satisfies Equation (13). We define the multiplicity vector of extended elements \( \gamma_{p''} \in \mathbb{Z}_{\geq 0}^{j} \) by

\[ \gamma_{p''} := \begin{cases} \hat{v}'_{p'}/f(p'') & \text{if } p'' = f(p') \\ 0 & \text{else.} \end{cases} \]

For the claim it remains to show that \( \gamma' \) satisfies the claimed properties. In particular, for every \( p'' \in P^{j-1} \) property (i) holds, since

\[ \alpha_{j-1} \beta_{j-1} \cdot \gamma_{p''} \leq \frac{1}{\Delta_t^2} \alpha_{j-1} \beta_{j-1} \cdot \gamma_{p''} \leq \lambda'_{p''}. \quad (14) \]

Every \( B \in \mathcal{B}^{j-1} \) has \( \|B\|_{\infty} \leq \Delta_{j-1} \). Hence, for every \( p \in P^j \) we have by Cramer’s rule and Hadamard’s bound that

\[ \|(B^p)^{-1} p\|_1 \leq d \cdot \|(B^p)^{-1} p\|_{\infty} \leq d(d\Delta_t)^d. \]

By the above, we can also bound

\[ \alpha_{j-1} \|\gamma'\|_1 = \alpha_{j-1} \sum_{p' \in P^{j-1}} \sum_{p' \in P^j} \gamma_{p'}((B^p)^{-1} p)p' \leq d(d\Delta_t)^d \alpha_{j-1} \|\gamma'\|_1. \]

By the definition of \( \hat{\gamma}' \), see Equation (5), we get \( \|\gamma'\|_1 = \|\hat{\gamma}'\|_1 \). Combined we can bound the size of the constructed multisets for each child \( \alpha_{j-1} \|\gamma'\|_1 \leq d(d\Delta_t)^d \alpha_{j-1} \|\gamma'\|_1 \), which is property (ii).

It remains to prove property (iii) \( \pi(\alpha_{j-1} \sum_{p' \in P^{j-1}} \gamma_{p'} p) = \alpha_{j-1} \sum_{p' \in P^{j-1}} \gamma_{p'} p \), which follows by the definitions of \( \hat{\gamma}' \) and \( \gamma' \). First, by the definition of \( \gamma' \) the following holds:

\[ \pi \left( \alpha_{j-1} \sum_{p' \in P^{j-1}} \gamma_{p'} p' \right) = \alpha_{j-1} \sum_{p' \in P^{j-1}} \gamma_{p'} p' = \alpha_{j-1} \sum_{p' \in P^{j-1}} \hat{\gamma}_{p'} p'. \quad (15) \]

Second, the definition of \( \hat{\gamma} \) yields property (iii),

\[ \alpha_{j-1} \sum_{p' \in P^{j-1}} \hat{\gamma}_{p'} p' = \alpha_{j-1} \sum_{p' \in P^{j-1}} \sum_{p' \in P^{j-1}} \gamma_{p'} (v(B^p)^{-1} p)p' = \alpha_{j-1} \sum_{p' \in P^{j-1}} \gamma_{p} B^p (B^p)^{-1} p = \alpha_{j-1} \sum_{p' \in P^{j-1}} \gamma_{p'} p. \quad (16) \]
Hence, combining the equalities (15) and (16) yields
\[
\pi \left( \alpha_{j-1} \sum_{\rho' \in P^{j-1}} Y^{\rho'}_{p'p''} \right) = \alpha_j \sum_{p \in P^j} Y^p_{p'}.
\]

We apply the claim iteratively from the root to the leaves on the inner vertices of \( T(A) \). In particular, we construct multiplicity vectors for every leaf \( v \in V \) that satisfy the properties. The constructed multiplicity vector \( Y^V \) of \( v \) is a submultiset of \( T_n(v) \), since by claim property (i) we have that
\[
Y^V = \alpha_0 \beta_0 \cdot Y^v \quad (i)
\]
and \( \lambda^v \) is defined as the multiplicity representation of multiset \( T_n(v) \).

Verifying the construction. Due to claim property (ii) the reconstruction of multisubsets grows at most
\[
\alpha_{j-1} \|Y^V\|_1 \leq d(d \Delta_t)^d \alpha_j \|Y^v\|_1
\]
from any vertex \( v \) to a child \( v' \). The scaling factor \( \alpha_t \) can be bounded by
\[
\alpha_t = \text{lcm}(1, \ldots, (d \Delta_t-1)^d)^t \leq 3^t(d \Delta_t-1)^d \leq 3^t(d \Delta)^{d^2+1} \leq 2^{d \Delta^2} \Delta^3
\]

since by the bound of Hanson [12] we have \( \text{lcm}(1, \ldots, k) \leq 3^k \). Hence, for a leaf \( v \) we have by induction that
\[
\|Y^v\|_1 = \alpha_0 \|Y^v\|_1 \quad (ii)
\]
\[
\leq d(d \Delta)^t \alpha_t \cdot \|Y^v\|_1 \quad (i)
\]
\[
= \alpha_0 \|Y^v\|_1 \leq 2^{d \Delta^2} \Delta^3
\]

To show that the multisubsets are valid for \( T(A) \) we construct a vector \( \hat{y} \in Z^{\text{col}(A)} \). The key observation here is property (iii) from the claim, which ensures that vertices connected in some subtree have equal sums in the indices of the root of the subtree. In each iteration, the sum for the parent defines the sum for every child in the respective indices. For an index \( k \in \text{col}(A) \) there exists \( v \in V \) of height \( j \leq t \) with \( k \in \hat{y} \). For precise indexing, let \( k' \) be the index of \( \hat{y}_k \) in the \( d_j \)-dimensional vector \( \pi^V(i, \hat{y}) \) for any \( i = n(v') \) and \( v' \) leaf in the subtree of \( v \). Define \( \hat{y}_k := \alpha_j \sum_{p \in P^j} Y^p_{p'k'} \). Next, we prove that the multisubsets are valid for \( T(A) \) regarding \( \hat{y} \). For a leaf \( v \in V \) with \( i = n(v) \), consider an index \( k \in \hat{y}_v \) for a vertex \( v' \in V \) of height \( j \leq t \) on the path from the root to \( v \). From claim property (iii) used inductively, we get
\[
\sum_{p \in P^j} Y^p_{p'k'} = \alpha_0 \sum_{p \in P^j} Y^p_{p'k'} \quad (ii)
\]
\[
= \alpha_0 \sum_{p \in P^j} Y^p_{p'k'} \quad \text{def.} = \hat{y}_k.
\]
As this holds for arbitrary indices on the path from the leaf to the root, we have that the constructed multisubsets are valid with respect to \( \hat{y} \), i.e.,
\[
\sum_{p \in P^j} Y^p_{p'} = \pi(i, \hat{y})
\]
as the vector \( \pi(i, \hat{y}) \) is composed of the indices of \( \hat{y} \) that lie on the path from the root to leaf \( v \). Due to the bounded size of the constructed multisubsets, the infinity norm of \( \hat{y} \) is bounded by
\[
\|\hat{y}\|_\infty \leq \Delta \cdot \max_{\text{leaf } v} \|Y^v\|_1 \quad (17)
\]
\[
\leq 2^{d \Delta^2} \Delta^3.
\]

The other case. Let us now turn to the case that there exists a vertex \( v \in V \) with \( \hat{y}^v > D_j = \alpha_j \beta_j \). We proceed very similarly to the above, but instead of reconstructing from the root, we reconstruct starting from \( v \) as we found a vector, that is 0, of high multiplicity. In this case, we can apply the claim to the unit vector \( Y^v := e_0 \), as it satisfies \( \alpha_j \beta_j e_0 \leq \lambda^v \). By the claim, we construct valid multisubsets for the subtree rooted in \( v \) of bounded size, similarly to the argumentation above. The
REFERENCES

[1] Matthias Aschenbrenner and Raymond Hemmecke. 2007. Finiteness theorems in stochastic integer programming. *Found. Comput. Math.* 7, 2 (2007), 183–227.

[2] John R. Birge and François Louveaux. 2011. *Introduction to Stochastic Programming*. Springer Science & Business Media.

[3] Gerard M. Campbell. 2011. A two-stage stochastic program for scheduling and allocating cross-trained workers. *J. Operat. Res. Soc.* 62, 6 (2011), 1038–1047.

[4] Lin Chen and Dániel Marx. 2018. Covering a tree with rooted subtrees - parameterized and approximation algorithms. In *Proc. SODA*. SIAM, 2801–2820.

[5] Jana Cslovjecsek, Friedrich Eisenbrand, Christoph Hunkenschröder, Lars Rohwedder, and Robert Weismantel. 2021. Block-structured integer and linear programming in strongly polynomial and near linear time. In *Proc. SODA*. SIAM, 1666–1681.

[6] Jana Cslovjecsek, Friedrich Eisenbrand, Michal Pilipczuk, Moritz Venzin, and Robert Weismantel. 2020. Efficient sequential and parallel algorithms for multistage stochastic integer programming using proximity. *CoRR* abs/2012.11742 (2020).

[7] Jana Cslovjecsek, Friedrich Eisenbrand, Michal Pilipczuk, Moritz Venzin, and Robert Weismantel. 2021. Efficient sequential and parallel algorithms for multistage stochastic integer programming using proximity. In *Proc. ESA*. 33:1–33:14.

[8] Friedrich Eisenbrand, Christoph Hunkenschröder, and Kim-Manuel Klein. 2018. Faster algorithms for integer programs with block structure. In *Proc. ICALP*. 49:1–49:13.

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

[10] Allen R. Ferguson and George B. Dantzig. 1956. The allocation of aircraft to routes—An example of linear programming under uncertain demand. *Manage. Sci.* 3, 1 (1956), 45–73.

[11] Michel Gendreau, Gilbert Laporte, and René Séguin. 1995. An exact algorithm for the vehicle routing problem with stochastic demands and customers. *Transport. Sci.* 29, 2 (1995), 143–155.

[12] Denis Hanson. 1972. On the product of the primes. *Can. Math. Bull.* 15, 1 (1972), 33–37.

[13] Raymond Hemmecke, Shmuel Onn, and Lyubov Romanchuk. 2013. n-Fold integer programming in cubic time. *Math. Program.* 137, 1–2 (2013), 325–341.

[14] Raymond Hemmecke and Rüdiger Schultz. 2003. Decomposition of test sets in stochastic integer programming. *Math. Program.* 94, 2–3 (2003), 323–341.

[15] Tsung-Wei Huang and Tsung-Yi Ho. 2011. A two-stage linear programming-based droplet routing algorithm for pin-constrained digital microfluidic biochips. *IEEE Trans. Comput.-aid. Des. Integr. Circ. Syst.* 30, 2 (2011), 215–228.

[16] Klaus Jansen, Kim-Manuel Klein, and Alexandra Lassota. 2021. The double exponential runtime is tight for 2-stage stochastic ILPs. In *Proc. IPCO*. 297–310.

[17] Klaus Jansen, Kim-Manuel Klein, Marten Maack, and Malin Rau. 2019. Empowering the configuration-IP - new PTAS results for scheduling with setups times. In *Proc. ITCS*.

[18] Klaus Jansen, Alexandra Lassota, and Marten Maack. 2020. Approximation algorithms for scheduling with class constraints. In *Proc. SPAA*. ACM, 349–357.

[19] Klaus Jansen, Alexandra Lassota, and Lars Rohwedder. 2020. Near-linear time algorithm for n-fold ILPs via color coding. *SIAM J. Discr. Math.* 34, 4 (2020), 2282–2299.

[20] Edward P. C. Kao and Maurice Queyranne. 1985. Budgeting costs of nursing in a hospital. *Manage. Sci.* 31, 5 (1985), 608–621.

[21] Kim-Manuel Klein. 2020. About the complexity of two-stage stochastic IPs. In *Proc. IPCO*. Springer, 252–265.

[22] Dusan Knop and Martin Koutecký. 2018. Scheduling meets n-fold integer programming. *J. Sched.* 21, 5 (2018), 493–503.

[23] Dusan Knop, Martin Koutecký, and Matthias Mnich. 2020. Combinatorial n-fold integer programming and applications. *Math. Program.* 184, 1 (2020), 1–34.

[24] Dusan Knop, Martin Koutecký, and Matthias Mnich. 2020. Voting and bribing in single-exponential time. *ACM Trans. Econ. Comput.* 8, 3 (2020), 12:1–12:28.
Collapsing the Tower—On the Complexity of Multistage Stochastic IPs

[25] Martin Koutecký, Asaf Levin, and Shmuel Onn. 2018. A parameterized strongly polynomial algorithm for block structured integer programs. In Proc. ICALP (LIPIcs). Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 85:1–85:14.

[26] Asaf Levin. 2022. Approximation schemes for the generalized extensible bin packing problem. Algorithmica 84, 2 (2022), 325–343.

[27] Vladimir I. Norkin, Yuri M. Ermoliev, and Andrzej Ruszczyński. 1998. On optimal allocation of indivisibles under uncertainty. Operat. Res. 46, 3 (1998), 381–395.

[28] Alexander Schrijver. 1998. Theory of Linear and Integer Programming. John Wiley & Sons.

[29] David B. Shmoys and Chaitanya Swamy. 2006. An approximation scheme for stochastic linear programming and its application to stochastic integer programs. Proc. JACM 53, 6 (2006), 978–1012.

[30] Guidong Zhu, Jonathan F. Bard, and Gang Yu. 2007. A two-stage stochastic programming approach for project planning with uncertain activity durations. J. Sched. 10, 3 (2007), 167–180.

Received 6 April 2022; revised 15 March 2023; accepted 29 May 2023