

Probabilistic analysis of dual decomposition on two-stage stochastic integer programs

Santanu S. Dey^{*1}, Marco Molinaro^{†2,3}, and Jingye Xu^{‡1}

¹Georgia Institute of Technology, Atlanta, GA, USA

²Microsoft Research, Redmond, WA, USA

³PUC-Rio, Rio de Janeiro, Brazil

April 25, 2026

Abstract

Two-stage stochastic integer programs provide a powerful framework for modeling decision-making under uncertainty, but they are notoriously difficult to solve at scale due to their high dimensionality and intrinsic nonconvexity. Decomposition-based algorithms such as Benders methods and Branch-and-Price (related dual decomposition methods) have become standard computational approaches for such problems and demonstrate excellent empirical performance in practice. Despite their widespread use, however, existing theoretical guarantees are almost exclusively based on worst-case analyses, which predict exponential convergence behavior in the problem dimension and fail to explain the strong performance observed in practice. In this paper, we present the first average-case analysis of Branch-and-Price for a broad class of two-stage stochastic binary integer programs. We study a stochastic-input model in which objective coefficients and constraint matrices are drawn at random and right-hand-side vectors scale with the decision dimension, while the number of constraints per scenario is fixed. Under this model, we prove that, with high probability, Branch-and-Price explores at most $n^{O(\log s)}$ nodes, yielding a quasi-polynomial bound on the size of the search tree in typical instances, where n denotes the decision dimension and s the number of scenarios. A key ingredient of our analysis is an average-case bound on the integrality gap of the natural linear programming (LP) relaxation. We show that this gap shrinks at rate $O\left(\frac{\log s \log^2 n}{n}\right)$ with high probability. This result is of independent interest, as it implies that the integrality gap grows only logarithmically with the number of scenarios on average. This helps explain why LP-based heuristics work well in practice, even in large-scale stochastic settings with an enormous number of scenarios, where LP relaxations with a large number of constraints are usually believed to be weak.

*Email: sdey30@gatech.edu

†Email: mmolinaro@microsoft.com

‡Email: jxu673@gatech.edu

1 Introduction

Consider the following problem:

$$\text{OPT}(\mathbf{b}) := \max_{\mathbf{x}} s \cdot \langle \mathbf{c}^{(0)}, \mathbf{x}^{(0)} \rangle + \sum_{i \in [s]} \langle \mathbf{c}^{(i)}, \mathbf{x}^{(i)} \rangle \quad (1a)$$

$$\text{s.t. } A^{(0)}\mathbf{x}^{(0)} + A^{(i)}\mathbf{x}^{(i)} \leq \mathbf{b}^{(i)}, \quad \forall i \in [s], \quad (1b)$$

$$\mathbf{x}^{(0)}, \mathbf{x}^{(i)} \in \{0, 1\}^n, \quad \forall i \in [s]. \quad (1c)$$

where $A^{(i)} \in \mathbb{Q}^{m \times n}$ and $\mathbf{c}^{(i)} \in \mathbb{Q}^n$ for all $i \in [s] \cup \{0\}$.

Problem (1) is a two-stage stochastic integer program, a widely used framework for modeling decision-making under uncertainty [15,16,19,20]. The vector $\mathbf{x}^{(0)}$ represents first-stage (here-and-now) decisions that must be fixed before the uncertainty is realized, while $\mathbf{x}^{(i)}$ denotes second-stage (recourse) decisions associated with scenario $i \in [s]$. Each scenario corresponds to a possible realization of the underlying uncertainty and is characterized by its right-hand side vector $\mathbf{b}^{(i)}$ and scenario-dependent constraint matrix $A^{(i)}$. In this paper, we focus on the setting in which all decision variables are binary and the first-stage and second-stage variables have the same dimension. These assumptions are made solely for notational simplicity and clarity of exposition. All results presented here extend in a straightforward manner to settings with heterogeneous variable dimensions, nonuniform scenario weights, and more general mixed-integer decision domains.

In practice, uncertainty is typically approximated by a finite set of scenarios obtained via sampling, scenario reduction, or discretization. Under this approximation, problem (1) can be viewed as a deterministic approximation of the original stochastic program. Although the dimension of (1) grows only linearly with the number of scenarios s , the associated computational burden often increases super-linearly. In the worst case, memory requirements, solution time, and solver performance deteriorate rapidly, and even evaluating a single candidate first-stage solution may require solving a large number of scenario-specific integer subproblems.

Decomposition methods are therefore particularly appealing for solving problem (1), as they exploit the inherent separability across scenarios [4,9,12,14,15,18]. By decomposing the problem into smaller, more tractable subproblems, these methods reduce memory usage, improve computational efficiency, and enable scalable solution approaches for large-scale instances. Two general decomposition paradigms are most commonly used in the literature for two-stage stochastic integer programs: *Benders decomposition* and *dual decomposition*. Benders decomposition is a cutting-plane approach that separates inequalities through value-function approximations of the second-stage problem, whereas dual decomposition reformulates the problem by introducing nonanticipativity constraints and relaxing these constraints via Lagrangian multipliers. For comprehensive surveys, we refer the reader to [18] for Benders decomposition and [7] for dual decomposition.

Despite their strong empirical performance, there remains a substantial gap between the practical success of decomposition methods and the current theoretical understanding of their convergence behavior for two-stage stochastic integer programs. Existing convergence guarantees are almost exclusively derived from worst-case analyses, which yield exponential complexity bounds of order $\Theta(2^n)$ [21,23]. Such bounds typically rely on the observation that there are at most 2^n possible first-stage integer solutions and that each iteration of a decomposition algorithm eliminates at least one suboptimal candidate. However, this pessimistic worst-case behavior is rarely observed in practice. Extensive computational studies indicate that decomposition methods often converge in only a small number of iterations, and in many cases even a single iteration suffices to identify a high-quality or optimal solution [14]. This pronounced discrepancy motivates the need for a refined, average-case analysis of the convergence behavior of decomposition methods.

In this paper, we focus on the dual decomposition approach. Because problem (1) is nonconvex,

there is generally a positive duality gap between (1) and the Lagrangian dual associated with its nonanticipativity reformulation. As a result, additional branching is required in practice to close this gap and recover an integral optimal solution. This leads to algorithmic frameworks commonly referred to as *Branch-and-Price*, *column generation*, or *Dantzig–Wolfe decomposition* [10, 17]. In this paper, we study the behavior of these algorithms.

To move beyond worst-case analysis, we adopt the common perspective of considering random input instances and focus on average-case performance. Specifically, we consider the following **stochastic-input model**:

1. All entries of $\mathbf{c}^{(i)}$ and $A^{(i)}$ are drawn independently and uniformly from $[0, 1]$, for all $i \in [s] \cup \{0\}$.
2. The right-hand-side vectors are given by $\mathbf{b}_r^{(i)} = \beta_r^{(i)} \cdot 2n$, where $\beta_r^{(i)} \in (\frac{1}{4}, \frac{1}{2})$ for all $i \in [s]$ and $r \in [m]$.

This random model has been considered in prior work (see [11, 13]) and is widely used to evaluate the performance of decomposition methods [1, 2]. Regarding the choice of the right-hand side, note that if $\beta_r^{(i)} > \frac{1}{2}$ for all i and r , then as $n \rightarrow \infty$ with m fixed, all binary solutions of (1) become feasible with high probability. In this regime, the problem admits the trivial optimal solution in which all decision variables equal one and thus becomes asymptotically uninformative. For technical reasons (see Section 4.3), we additionally impose the lower bound $\beta_r^{(i)} \geq \frac{1}{4}$, which is slightly stronger than assumptions used in [13]. Instances with very small values of $\beta_r^{(i)}$ are typically less interesting, as they admit only a limited number of feasible first-stage solutions. It is a folklore observation that the most challenging regime occurs when $\beta_r^{(i)}$ is close to $\frac{1}{2}$. We therefore believe that the proposed stochastic-input model is representative and well suited for evaluating the typical performance of decomposition-based algorithms.

In this work, we assume access to an oracle that can exactly solve the dual decomposition problem, and we measure the convergence behavior of Branch-and-Price by the number of nodes explored in the Branch-and-Price tree (Section 2 provides precise definitions). We emphasize that this oracle only requires an optimization oracle for each individual scenario subproblem of dimension n . Following prior work [5, 11, 13], we adopt a fixed-parameter regime in which the number of constraints in each scenario subproblem is fixed at m , while the decision dimension n and the number of scenarios s grow. Unlike previous studies, we do not fix the total number of constraints in problem (1); instead, the total number of constraints grows linearly with the number of scenarios. Our main result can be stated informally as follows.

Theorem 1.1 *In the stochastic-input model, for sufficiently large n , with probability at least $0.9999 - \frac{1}{n}$, the Branch-and-Price algorithm (see Section 2) solves (1) after exploring at most $n^{O_m(\log s)}$ nodes.*

Theorem 1.1 follows from the combination of two results established in the subsequent sections. The first result relates the size of the Branch-and-Price tree to the integrality gap of (1). The second result shows that this integrality gap is small with high probability under the stochastic-input model. Taken together, these results yield the stated bound on the number of nodes explored by the Branch-and-Price algorithm.

Lemma 1.2 *Consider the stochastic-input model, and suppose that, with probability at least $1 - \delta$, the integrality gap of problem (1) satisfies $O_m\left(\frac{s \log s \log^2 n}{n}\right)$.¹ Then, for sufficiently large n , with probability at least $1 - \delta - \frac{1}{n}$, the Branch-and-Price algorithm solves (1) after exploring at most $n^{O_m(\log s)}$ nodes.*

¹We use the notation $O_m(g(n, s))$ to denote a function $f(n, s, m)$ such that for every fixed $m \geq 0$, there exists a constant c such that $f(n, s, m) \leq c \cdot g(n, s)$ for all sufficiently large n, s .

Theorem 1.3 *In the stochastic-input model, for sufficiently large n , with probability at least 0.9999, the integrality gap of (1) satisfies $O_m\left(\frac{s \log s \log^2 n}{n}\right)$.*

To the best of our knowledge, this work provides the first theoretical average-case analysis of the Branch-and-Price algorithm, which has been a standard approach for solving large-scale integer programs since the 1990s [10,17]. Our result helps close the gap between the widespread practical success of Branch-and-Price and the lack of rigorous guarantees for its typical-case behavior, especially in high-dimensional and multi-scenario settings. Moreover, Theorem 1.3 extends the classical average-case integrality gap analysis of [13] from single-block integer programs to a much richer multi-block two-stage setting, while requiring only logarithmic growth in the number of scenarios s (after normalizing by $\frac{1}{s}$). This is of independent interest, as it shows that the LP relaxation remains tight even when the number of scenarios grows, a regime typically regarded as computationally prohibitive. In particular, it provides theoretical support for LP-relaxation-based decomposition methods [1, 8] by showing that their underlying relaxations continue to yield high-quality dual bounds in large-scale two-stage problems with many scenarios.

The remainder of the paper is devoted to proving Theorem 1.1 and is organized as follows. Section 2 presents a detailed description of the Branch-and-Price algorithm. Sections 3 and 4 prove Lemma 1.2 and Theorem 1.3, respectively.

2 Dual Decomposition and Branch-and-Price Algorithm

In this section, we provide a detailed description of a standard Branch-and-Price algorithm [3,10,17]. Similar to the standard Branch-and-Bound algorithm, which relies on LP relaxations, Branch-and-Price can be viewed as an enumeration procedure guided by a convex relaxation, where each node solves a dual decomposition problem (also known as column generation, and as “Lagrangian relaxation” in the stochastic programming literature).

Since in later sections we will also consider Lagrangian relaxations of linear programs, we adopt the term *dual decomposition relaxation* here to clearly distinguish between these two types of relaxations.

Dual decomposition relaxation. This continuous relaxation is obtained by convexifying each combination of the first-stage block and a single scenario block separately (whereas the original problem is equivalent to jointly convexifying the first-stage block together with all scenario blocks). More precisely, it is formally defined as

$$\begin{aligned} \text{DUAL} &:= \max_{\mathbf{x}} s \cdot \langle \mathbf{c}^{(0)}, \mathbf{x}^{(0)} \rangle + \sum_{i \in [s]} \langle \mathbf{c}^{(i)}, \mathbf{x}^{(i)} \rangle \\ &\text{s.t. } (\mathbf{x}^{(0)}, \mathbf{x}^{(i)}) \in \text{conv}(\mathcal{X}^{(i)}), \quad \forall i \in [s], \end{aligned} \tag{2}$$

where

$$\mathcal{X}^{(i)} := \left\{ (\mathbf{y}^{(0)}, \mathbf{y}^{(i)}) \left| \begin{array}{l} A^{(0)}\mathbf{y}^{(0)} + A^{(i)}\mathbf{y}^{(i)} \leq \mathbf{b}^{(i)}, \\ A^{(0)}\mathbf{y}^{(0)} \leq \mathbf{b}^{(j)}, \quad \forall j \in [s], \\ \mathbf{y}^{(0)}, \mathbf{y}^{(i)} \text{ binary.} \end{array} \right. \right\}.$$

We note that solving such a relaxation can be accomplished via standard column generation by solving multiple integer linear programs (ILPs) whose size is independent of the number of scenarios (more precisely, each has $2n$ variables and $2m$ linear constraints). Additionally, the Branch-and-Price algorithm only needs the optimal objective value of this relaxation, which can also be obtained by performing a

Lagrangian relaxation of the nonanticipativity constraints. This latter method, which also uses multiple solves of ILPs whose size is independent of s , is usually the one employed in practice. See, for example, [4] for more details.

Branch-and-Price. We consider the Branch-and-Price algorithm based on the dual decomposition relaxation. More precisely, we instantiate the elements of the Branch-and-Price framework as follows (see [10, 17] for a general description of Branch-and-Price):

1. **Node relaxation:** In each node N of the Branch-and-Price tree, we use the value $\text{DUAL}(N)$ of the dual decomposition relaxation (2) associated with that node (i.e., with the appropriate variables $\mathbf{x}^{(0)}$ fixed to 0/1).
2. **Branching rule:** Branch on any $\mathbf{x}^{(0)}$ variable that has not yet been fixed at the node; it does not need to be fractional.
3. **Node selection rule:** “Best-bound,” i.e., select the unpruned node with the largest computed dual value.
4. **Finding an incumbent solution:** Whenever we are at a node N where all variables $\mathbf{x}^{(0)}$ are fixed to 0/1 (call the fixing $\bar{\mathbf{x}}^{(0)}$), we compute a solution $\bar{\mathbf{x}}^{(1)}, \dots, \bar{\mathbf{x}}^{(s)}$ such that $\bar{\mathbf{x}}^{(0)}, \bar{\mathbf{x}}^{(1)}, \dots, \bar{\mathbf{x}}^{(s)}$ is feasible for the original problem and has value at least $\text{DUAL}(N)$.

3 Size of the Branch-and-Price Tree

In this section, we prove Lemma 1.2, namely, that as long as the integrality gap of problem (1) is $O_m\left(\frac{s \log s \log^2 n}{n}\right)$ with probability at least $1 - \delta$, then with probability at least $1 - \delta - \frac{1}{n}$, the Branch-and-Price algorithm solves (1) after exploring at most $n^{O_m(\log s)}$ nodes. (In Section 4, we prove this assumption on the integrality gap, thereby obtaining an unconditional bound on the size of the Branch-and-Price tree.)

The high-level idea for proving this result is as follows:

1. We consider the Lagrangian formulation (using an optimal dual solution) of the standard LP relaxation, and show that the primal optimal solution associated with each node in our Branch-and-Price (BP) tree has Lagrangian value at least OPT (Lemma 3.1). Intuitively, this holds because the “best-bound” rule guarantees that we never explore nodes with dual value below OPT, and the Lagrangian value is at least as high as our dual bound from (2).
2. We can associate with each node in our BP tree a **binary** vector with Lagrangian value at least OPT, and at most n internal nodes are associated with the same vector. This implies an upper bound on the size of our BP tree based on the number of such binary vectors with high Lagrangian value (Lemma 3.2).

This is accomplished by appropriately rounding the primal optimal solution at each node, using the fact that the Lagrangian value is linear once the Lagrangian dual has been fixed.

3. We prove that, with high probability, there are only $n^{O_m(\log s)}$ binary vectors with such high Lagrangian value, as long as the integrality gap is small.

While the overall argument is similar to that of [11], the key difference is that the original bound is exponential in the number of constraints. In our setting, this number is $m \cdot s$, which would lead to an

exponential dependence on the number of scenarios. By contrast, our bound scales as $n^{O_m(\log s)}$. To obtain this improved dependence on s , we work only with the projection onto the first-stage variables $\mathbf{x}^{(0)}$, that is, we essentially restrict attention to the matrix $A^{(0)}$, which has only m rows. This approach exploits the fact that the relaxation (2) convexifies each scenario jointly with the first-stage variables, and is therefore stronger than the LP relaxation considered in [11].

To formalize this argument, consider the standard LP relaxation of the full problem:

$$\text{OPT}_{\text{LP}} := \max s \cdot \langle \mathbf{c}^{(0)}, \mathbf{x}^{(0)} \rangle + \sum_{i \in [s]} \langle \mathbf{c}^{(i)}, \mathbf{x}^{(i)} \rangle \quad (3a)$$

$$\text{s.t. } A^{(0)}\mathbf{x}^{(0)} + A^{(i)}\mathbf{x}^{(i)} \leq \mathbf{b}^{(i)}, \quad \forall i \in [s], \quad (3b)$$

$$\mathbf{0} \leq \mathbf{x} \leq \mathbf{1}. \quad (3c)$$

Define the integrality gap by

$$\Delta := \text{OPT}_{\text{LP}} - \text{OPT}.$$

Consider the Lagrangian relaxation of this linear program:

$$\begin{aligned} \text{OPT}_{\text{LP}} &= \min_{\mu \geq 0} \max_{0 \leq \mathbf{x} \leq 1} \left(s \cdot \langle \mathbf{c}^{(0)}, \mathbf{x}^{(0)} \rangle + \sum_{r \in [s]} \langle \mathbf{c}^{(r)}, \mathbf{x}^{(r)} \rangle \right) + \left(\sum_{r \in [s]} \langle \mu^{(r)}, \mathbf{b}^{(r)} - A^{(0)}\mathbf{x}^{(0)} - A^{(r)}\mathbf{x}^{(r)} \rangle \right) \\ &= \min_{\mu \geq 0} \max_{0 \leq \mathbf{x} \leq 1} s \cdot \left\langle \mathbf{c}^{(0)} - (A^{(0)})^\top \left(\frac{1}{s} \sum_r \mu^{(r)} \right), \mathbf{x}^{(0)} \right\rangle + \sum_{r \in [s]} \langle \mathbf{c}^{(r)} - (A^{(r)})^\top \mu^{(r)}, \mathbf{x}^{(r)} \rangle + \sum_{r \in [s]} \langle \mu^{(r)}, \mathbf{b}^{(r)} \rangle. \end{aligned} \quad (4)$$

Let (\mathbf{x}_*, μ_*) be the saddle point of this Lagrangian. To simplify the notation, let $\bar{\mu}_* := \frac{1}{s} \sum_{i \in [s]} \mu_*^{(i)}$. Given any first-stage point $\mathbf{x}^{(0)}$, let

$$\Gamma(\mathbf{x}^{(0)}) := s \cdot \left\langle \mathbf{c}^{(0)} - (A^{(0)})^\top \bar{\mu}_*, \mathbf{x}^{(0)} \right\rangle + \underbrace{\sum_{r \in [s]} \langle \mathbf{c}^{(r)} - (A^{(r)})^\top \mu_*^{(r)}, \mathbf{x}_*^{(r)} \rangle + \sum_{r \in [s]} \langle \mu_*^{(r)}, \mathbf{b}^{(r)} \rangle}_{\text{constant}}$$

be the ‘‘Lagrangian value’’ of this point; note that we are using the optimal dual μ_* and also the optimal primal $\mathbf{x}_*^{(r)}$ for all the scenarios. Define the set of first-stage binary solutions with Γ -value at least OPT as (recall that OPT is the integer optimum)

$$\Omega := \{\mathbf{x}^{(0)} \in \{0, 1\}^n : \Gamma(\mathbf{x}^{(0)}) \geq \text{OPT}\}.$$

The first claim is that for every node in the Branch-and-Price tree, the primal optimal solution for the relaxation at that node has Γ -value at least OPT .

Lemma 3.1 *Consider any internal node in the Branch-and-Price tree, and let $\bar{\mathbf{x}}^{(0)}, \bar{\mathbf{x}}^{(1)}, \dots, \bar{\mathbf{x}}^{(s)}$ be the optimal primal solution of the relaxation (2) at this node. Then $\Gamma(\bar{\mathbf{x}}^{(0)}) \geq \text{OPT}$.*

PROOF: Let $\text{DUAL}(N)$ denote the exact optimal value of the relaxation (2) associated with node N in the BP tree. During the execution of BP, the value $\text{DUAL}(N)$ is used as the node-selection criterion under the best-bound rule.

We claim that every internal node N explored by BP satisfies $\text{DUAL}(N) \geq \text{OPT}$. By contradiction, suppose there exists a node N with $\text{DUAL}(N) < \text{OPT}$ that is selected for branching. By the best-bound rule, this implies that for every other open node N' in the BP tree, $\text{DUAL}(N') \leq \text{DUAL}(N) <$

OPT. On the other hand, there exists at least one node N^* (open or pruned) whose relaxation contains the globally optimal integral solution of the original problem. For such a node, we have $\text{OPT} \leq \text{DUAL}(N^*)$. Therefore, N^* cannot be an open node at the time N is selected, since all open nodes have $\text{DUAL}(\cdot) < \text{OPT}$. Hence, N^* must already be pruned. If N^* was pruned by integrality, then an incumbent solution of value OPT has already been found. If N^* was pruned by bounds, then an incumbent of value at least OPT must have been identified prior to pruning. In either case, the current incumbent value is OPT. However, since $\text{DUAL}(N) < \text{OPT}$, node N would be pruned by the bounding rule and thus could not be selected for branching, contradicting our assumption. We conclude that every internal node of the BP tree satisfies $\text{DUAL}(N) \geq \text{OPT}$.

To conclude the proof, we show that for any node N in the BP tree, the optimal primal solution $\bar{\mathbf{x}}^{(0)}, \bar{\mathbf{x}}^{(1)}, \dots, \bar{\mathbf{x}}^{(s)}$ of its relaxation (2) has Γ -value at least $\text{DUAL}(N)$, namely,

$$\Gamma(\bar{\mathbf{x}}^{(0)}) \geq s \cdot \langle \mathbf{c}^{(0)}, \bar{\mathbf{x}}^{(0)} \rangle + \sum_{r \in [s]} \langle \mathbf{c}^{(r)}, \bar{\mathbf{x}}^{(r)} \rangle. \quad (5)$$

To prove this, we expand the definition of Γ and use the fact that $\mathbf{x}_*^{(r)}$ maximizes $\langle \mathbf{c}^{(r)} - (A^{(r)})^\top \mu_*^{(r)}, \mathbf{x}^{(r)} \rangle$ to obtain

$$\begin{aligned} \Gamma(\bar{\mathbf{x}}^{(0)}) &\geq s \cdot \langle \mathbf{c}^{(0)} - (A^{(0)})^\top \bar{\mu}_*, \bar{\mathbf{x}}^{(0)} \rangle + \sum_{r \in [s]} \langle \mathbf{c}^{(r)} - (A^{(r)})^\top \mu_*^{(r)}, \bar{\mathbf{x}}^{(r)} \rangle + \sum_{r \in [s]} \langle \mu_*^{(r)}, \mathbf{b}^{(r)} \rangle \\ &= s \cdot \langle \mathbf{c}^{(0)}, \bar{\mathbf{x}}^{(0)} \rangle + \sum_{r \in [s]} \langle \mathbf{c}^{(r)}, \bar{\mathbf{x}}^{(r)} \rangle - \sum_{r \in [s]} \langle \mu_*^{(r)}, A^{(0)} \bar{\mathbf{x}}^{(0)} + A^{(r)} \bar{\mathbf{x}}^{(r)} \rangle + \sum_{r \in [s]} \langle \mu_*^{(r)}, \mathbf{b}^{(r)} \rangle. \end{aligned} \quad (6)$$

Since the solution $\bar{\mathbf{x}}^{(0)}, \bar{\mathbf{x}}^{(1)}, \dots, \bar{\mathbf{x}}^{(s)}$ is feasible for the original constraints, i.e.,

$$A^{(0)} \bar{\mathbf{x}}^{(0)} + A^{(r)} \bar{\mathbf{x}}^{(r)} \leq \mathbf{b}^{(r)} \quad \text{for all } r \in [s],$$

and $\mu_*^{(r)} \geq 0$, the last two terms of (6) sum to a nonnegative value. Therefore,

$$\Gamma(\bar{\mathbf{x}}^{(0)}) \geq s \cdot \langle \mathbf{c}^{(0)}, \bar{\mathbf{x}}^{(0)} \rangle + \sum_{r \in [s]} \langle \mathbf{c}^{(r)}, \bar{\mathbf{x}}^{(r)} \rangle,$$

which proves (5). This completes the proof of the lemma. \square

Given this result, the point $\bar{\mathbf{x}}^{(0)}$ in the lemma would belong to the set Ω if $\bar{\mathbf{x}}^{(0)}$ had only 0/1 coordinates. However, this point is only an optimal solution of the relaxation (2) (at some node) and thus may have fractional coordinates. Since $\Gamma(\cdot)$ is a linear function, we can round this point to a 0/1 point with at least as large a Γ -value. More precisely, for a node N of our BP tree, define $\mathbf{x}_N^{(0)}$ as the argmax

$$\mathbf{x}_N^{(0)} := \operatorname{argmax} \left\{ \Gamma(\mathbf{x}^{(0)}) : \mathbf{x}^{(0)} \in \{0, 1\}^n \text{ and } \mathbf{x}^{(0)} \text{ satisfies the variable fixings of the BP tree at } N \right\}.$$

We now have that these solutions $\mathbf{x}_N^{(0)}$ belong to Ω , and using this we can bound the size of our BP tree by the size of Ω .

Lemma 3.2 *For every internal node N in the BP tree, $\mathbf{x}_N^{(0)} \in \Omega$. Moreover, for each $\mathbf{x}^{(0)} \in \Omega$, there are at most n internal nodes N in the BP tree such that $\mathbf{x}_N^{(0)} = \mathbf{x}^{(0)}$. In particular, the size of the BP tree is at most $2n \cdot |\Omega| + 1$.*

PROOF: For the first part: for an internal node N in the BP tree, the solution $\bar{\mathbf{x}}^{(0)}$ from Lemma 3.1 is feasible for the argmax defining $\mathbf{x}_N^{(0)}$, which implies $\Gamma(\mathbf{x}_N^{(0)}) \geq \Gamma(\bar{\mathbf{x}}^{(0)}) \geq \text{OPT}$, and so $\mathbf{x}_N^{(0)} \in \Omega$.

For the second property: if two internal nodes N_1 and N_2 are not on the same path of the BP tree, then by looking at the common ancestor of these nodes we observe that there is some coordinate $\mathbf{x}_j^{(0)}$ that is fixed to 0 in N_1 and 1 in N_2 , or vice versa; this implies that their points $\mathbf{x}_{N_1}^{(0)}$ and $\mathbf{x}_{N_2}^{(0)}$ are different (in coordinate j). Thus, all the internal nodes N whose solutions $\mathbf{x}_N^{(0)}$ map to the same point $\mathbf{x}^{(0)}$ of Ω must lie along a single path, and thus there are at most n of them (recall that our Branch-and-Price algorithm only branches on the n first-stage variables $\mathbf{x}^{(0)}$).

This implies that there are at most $n \cdot |\Omega|$ internal nodes in the BP tree, which further implies at most $2n \cdot |\Omega| + 1$ total nodes (since it is a binary tree). This concludes the proof. \square

In light of this result, in order to bound the size of the BP tree, it suffices to bound the number of points in Ω .

Lemma 3.3 *Suppose that with probability at least $1 - \delta$, $\Delta \leq O_m(\frac{s \log s \log^2 n}{n})$. Then with probability at least $1 - \frac{1}{n} - \delta$, we have $|\Omega| \leq n^{O_m(\log s)}$.*

Lemmas 3.2 and 3.3 thus imply the desired Lemma 1.2.

3.1 Proof of Lemma 3.3

Recall that (\mathbf{x}_*, μ_*) is the saddle point of the Lagrangian (4), and that $\bar{\mu}_* = \frac{1}{s} \sum_r \mu_*^{(r)}$.

Recalling the definition of Ω , we would like to bound the number of 0/1 points $\mathbf{x}^{(0)}$ such that $\Gamma(\mathbf{x}^{(0)}) \geq \text{OPT}$. For that, consider the following hyperplane defined by the dual $\bar{\mu}_*$:

$$H := \{y \in \mathbb{R}^{m+1} : \langle (1, -\bar{\mu}_*), y \rangle = 0\}.$$

The relevance of this hyperplane is that it transfers our task into a geometric one, since, as we show, $\Gamma(\mathbf{x}^{(0)})$ can be related to the distance from the first-stage columns $(c_j^{(0)}, \mathbf{a}_j^{(0)})$ to the hyperplane H for the coordinates j where $\mathbf{x}^{(0)}$ and $\mathbf{x}_*^{(0)}$ disagree. Let $d(y, H)$ denote the Euclidean distance between a point y and the hyperplane H .

Lemma 3.4 *For every $\mathbf{x}^{(0)} \in \{0, 1\}^n$ we have*

$$\Gamma(\mathbf{x}^{(0)}) \leq \text{OPT} + \Delta - s \cdot \sum_j d\left((c_j^{(0)}, \mathbf{a}_j^{(0)}), H\right) \cdot \mathbf{1}(x_j^{(0)} \neq x_{*,j}^{(0)}).$$

Consequently, if $\mathbf{x}^{(0)}$ belongs to Ω , then

$$\sum_j d\left((c_j^{(0)}, \mathbf{a}_j^{(0)}), H\right) \cdot \mathbf{1}(x_j^{(0)} \neq x_{*,j}^{(0)}) \leq \frac{\Delta}{s}.$$

PROOF: The ‘‘Consequently’’ part follows directly from the definition of Ω , so we only prove the first part of the lemma.

Since (\mathbf{x}_*, μ_*) is a saddle point for the Lagrangian (4), by strong duality it achieves the value of OPT_{LP} , and so

$$\begin{aligned} \text{OPT}_{\text{LP}} - \Gamma(\mathbf{x}^{(0)}) &= s \cdot \left\langle \mathbf{c}^{(0)} - (A^{(0)})^\top \bar{\mu}_*, \mathbf{x}_*^{(0)} - \mathbf{x}^{(0)} \right\rangle \\ &= s \cdot \sum_{j=1}^n \left(\mathbf{c}^{(0)} - (A^{(0)})^\top \bar{\mu}_* \right)_j \cdot (x_{*,j}^{(0)} - x_j^{(0)}). \end{aligned} \quad (7)$$

Moreover, using the fact that $\mathbf{x}_*^{(0)}$ is a saddle point and $\mathbf{x}^{(0)}$ is a 0/1 vector, we claim that for every j we have

$$\left(\mathbf{c}^{(0)} - (A^{(0)})^\top \bar{\boldsymbol{\mu}}_*\right)_j \cdot (x_{*,j}^{(0)} - x_j^{(0)}) = \left| \left(\mathbf{c}^{(0)} - (A^{(0)})^\top \bar{\boldsymbol{\mu}}_*\right)_j \right| \cdot \mathbf{1}(x_j^0 \neq x_{*,j}^0). \quad (8)$$

This is because if $\left(\mathbf{c}^{(0)} - (A^{(0)})^\top \bar{\boldsymbol{\mu}}_*\right)_j > 0$, then $x_{*,j}^{(0)} = 1$ (since it is a saddle point) and so $(x_{*,j}^{(0)} - x_j^{(0)}) = \mathbf{1}(x_j^0 \neq x_{*,j}^0)$, whereas if $\left(\mathbf{c}^{(0)} - (A^{(0)})^\top \bar{\boldsymbol{\mu}}_*\right)_j < 0$, then $x_{*,j}^{(0)} = 0$ and $-(x_{*,j}^{(0)} - x_j^{(0)}) = \mathbf{1}(x_j^0 \neq x_{*,j}^0)$.

Finally, we claim that

$$d\left(\left(c_j^{(0)}, \mathbf{a}_j^{(0)}\right), H\right) \leq \left| \left(\mathbf{c}^{(0)} - (A^{(0)})^\top \bar{\boldsymbol{\mu}}_*\right)_j \right|. \quad (9)$$

This is because the point $\left(\langle \bar{\boldsymbol{\mu}}_*, \mathbf{a}_j^{(0)} \rangle, \mathbf{a}_j^{(0)}\right)$ belongs to H , and so it upper-bounds the distance:

$$\begin{aligned} d\left(\left(c_j^{(0)}, \mathbf{a}_j^{(0)}\right), H\right) &\leq \left\| \left(c_j^{(0)}, \mathbf{a}_j^{(0)}\right) - \left(\langle \bar{\boldsymbol{\mu}}_*, \mathbf{a}_j^{(0)} \rangle, \mathbf{a}_j^{(0)}\right) \right\|_2 \\ &= \left| c_j^{(0)} - \langle \bar{\boldsymbol{\mu}}_*, \mathbf{a}_j^{(0)} \rangle \right| = \left| \left(\mathbf{c}^{(0)} - (A^{(0)})^\top \bar{\boldsymbol{\mu}}_*\right)_j \right|. \end{aligned}$$

Combining (8) and (9) and applying them to (7), we get

$$\text{OPT}_{\text{LP}} - \Gamma(\mathbf{x}^{(0)}) \geq s \cdot \sum_{j=1}^n d\left(\left(c_j^{(0)}, \mathbf{a}_j^{(0)}\right), H\right) \cdot \mathbf{1}(x_j^0 \neq x_{*,j}^0).$$

Using the definition of the integrality gap $\Delta = \text{OPT}_{\text{LP}} - \text{OPT}$ and rearranging the terms then gives the claim of the lemma. \square

Roughly speaking, the previous lemma means that if $\mathbf{x}^{(0)} \in \Omega$, then the vector $\mathbf{x}^{(0)}$ and the saddle point $\mathbf{x}_*^{(0)}$ can differ only on coordinates j for which the column $\left(c_j^{(0)}, \mathbf{a}_j^{(0)}\right)$ is close to the hyperplane H . We now show that, with high probability, there are only “a few” of these columns; this in turn will imply that there are only “a few” possibilities for the vectors in Ω to differ from $\mathbf{x}_*^{(0)}$, which upper-bounds the size of Ω as desired.

To make this precise, we define “buckets” of columns based on their distances to H : for $\ell \geq 1$ define

$$J_\ell := \left\{ j : d\left(\left(c_j^{(0)}, \mathbf{a}_j^{(0)}\right), H\right) \text{ is in the interval } \left(\frac{\log n}{n} 2^\ell, \frac{\log n}{n} 2^{\ell+1}\right] \right\},$$

and define $J_{\text{rem}} := [n] \setminus \bigcup_{\ell \geq 1} J_\ell$ as the remaining columns (i.e., those with distance at most $\frac{\log n}{n} \cdot 2$). The next lemma establishes the connection between the number of column elements in these buckets and the size of Ω . The proof is the same as that of Lemma 5 of [11], but we reproduce it for completeness. We use the notation $\binom{u}{\leq v} := \binom{u}{0} + \dots + \binom{u}{v}$ for positive integers $u \geq v$.

Lemma 3.5 *We have the following upper bound:*

$$|\Omega| \leq 2^{|J_{\text{rem}}|} \cdot \prod_{\ell=1}^{\log C} \binom{|J_\ell|}{\leq C/2^\ell},$$

where $C := \frac{n}{\log n} \cdot \frac{\Delta}{s}$.

PROOF: First, observe that for every column j in $\bigcup_{\ell \geq 1} J_\ell$, the coordinate $x_{*,j}^{(0)}$ has value 0 or 1: these columns have strictly positive distance to H , hence from (9) they have nonzero reduced costs $\left(\mathbf{c}^{(0)} - (A^{(0)})^\top \bar{\boldsymbol{\mu}}_*\right)_j$, and so the saddle-point solution $x_{*,j}^{(0)}$ is at the boundary of its domain $[0, 1]$.

Next, notice that every point in Ω can be thought of as being created by starting with the vector $\mathbf{x}_*^{(0)}$ and then changing some of its coordinates, and because of Lemma 3.4 we:

- cannot change the value of $\mathbf{x}_*^{(0)}$ in any coordinate j in a bucket J_ℓ with $\ell > \log C$;
- can only flip the value of $\mathbf{x}_*^{(0)}$ in at most $\frac{C}{2^\ell}$ of the coordinates $j \in J_\ell$, for each $\ell = 1, \dots, \log C$ (recall that both $x_{*,j}^{(0)}$ and $x_j^{(0)}$ take only the values 0 or 1);
- may set a new arbitrary 0/1 value for, in principle, all coordinates in J_{rem} .

Since there are at most $2^{|J_{\text{rem}}|} \cdot \prod_{\ell=1}^{\log C} \binom{|J_\ell|}{\leq C/2^\ell}$ possibilities in this process, we have the desired upper bound. \square

Since both the columns $(c_j^{(0)}, \mathbf{a}_j^{(0)})$ and the hyperplane are random objects that depend on the entries of the input problem, the sizes of the sets J_ℓ are also random. In order to control these sizes with high probability, we will use the following technical lemma (Theorem 4 of [11]); let \mathbb{S}^{k-1} denote the unit sphere in \mathbb{R}^k .

Lemma 3.6 (Theorem 4 of [11]) *For $u \in \mathbb{S}^{k-1}$ and $w \geq 0$, define the slab of normal u and width w as*

$$S_{u,w} := \left\{ y \in \mathbb{R}^k : \langle u, y \rangle \in [-w, w] \right\}.$$

Let Y^1, \dots, Y^n be independent random vectors uniformly distributed in the cube $[0, 1]^k$. For $n \geq k$, then with probability at least $1 - \frac{1}{n}$, we have that for all $u \in \mathbb{S}^{k-1}$ and $w \geq \frac{\log n}{n}$, at most $60nw$ of the Y^j 's belong to $S_{u,w}$.²

Notice that if a column $(c_j^{(0)}, \mathbf{a}_j^{(0)})$ is such that j belongs to J_ℓ , then this column belongs to the $(m+1)$ -dimensional slab $\left\{ y : d(y, H) \leq \frac{\log n}{n} 2^{\ell+1} \right\}$ of width $\frac{\log n}{n} 2^{\ell+1}$; thus, the previous lemma can be used to upper-bound the size $|J_\ell|$ (and similarly for $|J_{\text{rem}}|$). Notice that we do not even need to take a union bound over the ℓ 's, since the previous lemma controls their sizes simultaneously. This gives the following:

Corollary 3.1 *With probability at least $1 - \frac{1}{n}$, we have simultaneously*

$$\begin{aligned} |J_{\text{rem}}| &\leq 120(m+1) \log n, \\ |J_\ell| &\leq 60(m+1) 2^{\ell+1} \log n, \quad \forall \ell \in [\log n - 1]. \end{aligned}$$

Now we are ready to prove Lemma 3.3.

PROOF OF LEMMA 3.3: Since we assumed the integrality gap satisfies $\Delta \leq O_m\left(\frac{s \log s \log^2 n}{n}\right)$ with probability at least $1 - \delta$, by taking a union bound, with probability at least $1 - \frac{1}{n} - \delta$ both this event and the upper bounds on $|J_\ell|$ and $|J_{\text{rem}}|$ from Corollary 3.1 hold. We show that in this case $|\Omega| \leq n^{O_m(\log s)}$, proving the lemma.

Under this event, we have $C := \frac{n}{\log n} \cdot \frac{\Delta}{s} \leq f(m) \cdot \log s \log n$ for some increasing function $f(\cdot) \geq 1$. Let $\ell_0 \geq 0$ be a sufficiently large integer of order $\frac{\log(O_m(\log s))}{2}$ but satisfying $60(m+1)2^{2\ell_0+1} \log n \geq 4C$ (which exists by the previous upper bound on C). We will use the following standard estimate:

²We note that the assumption $n \geq k$ is without loss of generality, since the result is trivially true otherwise: the assumption $w \geq \frac{\log n}{n}$ implies that the final upper bound is $60nw \geq 60 \log n \cdot k > k$, making the statement vacuous when $n \leq k$.

$\binom{a}{\leq b} \leq \left(\frac{4a}{b}\right)^b$ whenever $a \geq 4b$ (our definition of ℓ_0 ensures that this requirement holds in our application of the estimate). Then we have the following upper bound:

$$\begin{aligned} \prod_{\ell=\ell_0}^{\log C} \binom{|J_\ell|}{\leq C/2^\ell} &\leq \prod_{\ell=\ell_0}^{\log C} \left(\frac{240(m+1)2^{\ell+1} \log n}{(f(m) \cdot \log s \log n)/2^\ell} \right)^{C/2^\ell} \leq \prod_{\ell=\ell_0}^{\log C} \left(\frac{O_m(1) \cdot 2^{2\ell}}{\log s} \right)^{C/2^\ell} \\ &\leq \left(\frac{O_m(1)}{\log s} \right)^{C \sum_{\ell \geq 1} \frac{1}{2^\ell}} \cdot 2^{2C \sum_{\ell \geq 1} \frac{\ell}{2^\ell}} \\ &\leq \left(\frac{O_m(1)}{\log s} \right)^{O(C)} \cdot 2^{O_m(C)} \leq n^{O_m(\log s)}. \end{aligned}$$

For the terms with $\ell < \ell_0$, we can use the crude estimate $\binom{a}{b} \leq 2^a$ to get

$$\prod_{\ell=1}^{\ell_0-1} \binom{|J_\ell|}{\leq C/2^\ell} \leq \prod_{\ell=1}^{\ell_0-1} 2^{|J_\ell|} \leq 2^{O_m(1) \cdot \ell_0 \cdot 2^{\ell_0} \log n} \leq 2^{O_m(1) \cdot \log \log s \cdot \sqrt{\log s} \cdot \log n} \leq n^{O_m(\log s)}.$$

Finally, directly from Corollary 3.1 we have $|J_{\text{rem}}| \leq O_m(\log n)$. Employing these bounds in Lemma 3.5 we obtain

$$|\Omega| \leq 2^{|J_{\text{rem}}|} \cdot \prod_{\ell=1}^{\log C} \binom{|J_\ell|}{\leq C/2^\ell} \leq n^{O_m(\log s)}.$$

This concludes the proof of Lemma 3.3. \square

4 Integrality Gap of Random Two-stage Integer Programs

In this section, we prove Theorem 1.3. Our approach extends the probabilistic analysis of [13], which was developed for single-block packing integer programs, to a broader class of multi-block packing integer programs with a two-stage structure. The main technical difficulty, as well as a detailed comparison with the argument of [13], are deferred to Section 4.1, where the necessary background has been introduced.

Recall that the natural LP relaxation of (1), parameterized by the right-hand side vector \mathbf{b} , is given by

$$\text{OPT}_{\text{LP}}(\mathbf{b}) := \max s \langle \mathbf{c}^{(0)}, \mathbf{x}^{(0)} \rangle + \sum_{i \in [s]} \langle \mathbf{c}^{(i)}, \mathbf{x}^{(i)} \rangle \quad (10a)$$

$$\text{s.t. } A^{(0)} \mathbf{x}^{(0)} + A^{(i)} \mathbf{x}^{(i)} \leq \mathbf{b}^{(i)}, \quad \forall i \in [s], \quad (10b)$$

$$\mathbf{0} \leq \mathbf{x} \leq \mathbf{1}. \quad (10c)$$

and Δ denotes the (additive) integrality gap of this relaxation, namely

$$\Delta := \text{OPT}_{\text{LP}}(\mathbf{b}) - \text{OPT}(\mathbf{b}).$$

The formal statement of Theorem 1.3 is as follows:

Theorem 4.1 *Let $\beta_{\min} := \min_{i \in [s], r \in [m]} \beta_r^{(i)}$ and set $\alpha := \beta_{\min} - \frac{1}{4}$. Under this stochastic model, and under the assumptions that $n \geq 2^{m+4}(m+2) \log(2mn)$, $2\sqrt{n \ln snm} + 6\alpha m \log n < n \cdot (\beta_{\min} - \frac{1}{4})$, $\frac{2\sqrt{3m}}{\alpha} \leq$*

$(m \log n)^{1/4}$, and n is at least a sufficiently large constant, there exists a constant $\pi \in (0, 1)$ that only depends on m, β such that

$$\mathbb{P}\left(\Delta \leq O_m\left(\frac{sf \log^2 n}{n}\right)\right) \geq 1 - \frac{2s+3}{n^2} - (s+1)\pi^f$$

for all $f \geq \max\left\{\frac{\alpha^{2m+6} \log(2mn)}{\log n}, \frac{2^{2m+8}}{\alpha^m}\right\}$. By choosing $f = \sigma \log s$ for some sufficiently large constant σ (depends on m, β) and for sufficiently large n , this implies that

$$\mathbb{P}\left(\Delta \leq O_m\left(\frac{s \log s \log^2 n}{n}\right)\right) \geq 0.9999.$$

The full proof of this theorem is rather involved and we sketch the main proof ideas in Section 4.1.

4.1 Main Steps of Theorem 4.1

Since the proof of Theorem 4.1 consists of several technical components, we first present the key lemmas and explain how they together imply the theorem. Each lemma is then proved in a separate subsection.

To bound the additive integrality gap $\Delta = \text{OPT}_{\text{LP}}(\mathbf{b}) - \text{OPT}(\mathbf{b})$, we first derive a bound on how the objective value changes when moving from an optimal LP solution to an integral one (e.g., via rounding). In later sections, we show how to construct such a desired rounding. To get a hold of the LP relaxation, it will be useful to consider its saddle-point (min-max) formulation similar to the previous argument. Let

$$L_{\mathbf{b}}^{\text{std}}(\mathbf{x}, \mu) := s \cdot \langle \mathbf{c}^{(0)}, \mathbf{x}^{(0)} \rangle + \sum_{i \in [s]} \langle \mathbf{c}^{(i)}, \mathbf{x}^{(i)} \rangle + \sum_{i \in [s]} \langle \mu^{(i)}, \mathbf{b}^{(i)} - A^{(0)} \mathbf{x}^{(0)} - A^{(i)} \mathbf{x}^{(i)} \rangle \quad (11)$$

denote the objective of this Lagrangian relaxation, so the LP value is $\text{OPT}_{\text{LP}} = \max_{0 \leq \mathbf{x} \leq \mathbf{1}} \min_{\mu \geq 0} L_{\mathbf{b}}^{\text{std}}(\mathbf{x}, \mu)$.

We use $\beta_{\min} := \min_{i \in [s], r \in [m]} \beta_r^{(i)}$, and set $\alpha := \beta_{\min} - \frac{1}{4}$. With a lot of hindsight, define

$$d := \frac{3\alpha m \log n}{2} \cdot \left(1 + \frac{1}{\sqrt{3} \cdot n^2}\right); \quad (12)$$

We will actually consider optimal primal/dual solution for the LP with modified RHS $\mathbf{b}_{\diamond} := \mathbf{b} - 2d\mathbf{1}$. This extra slack will be necessary to ensure feasibility when we perform our rounding in Lemma 4.2, in Section 4.6. The following lemma is used to control integrality gap Δ . Here, $\mathbf{a}_j^{(i)}$ denotes the j -th column of the matrix $A^{(i)}$ and $\mathbf{x}_j^{(i)}$ denotes the j -th entry of the vector $\mathbf{x}^{(i)}$.

Lemma 4.1 *Let $(\mathbf{x}_{\diamond}, \mu_{\diamond})$ be saddle-point for $L_{\mathbf{b}_{\diamond}}^{\text{std}}(\mathbf{x}, \mu)$ (notice the modified right-hand side \mathbf{b}_{\diamond}); define $\mu^{(0)} := \frac{1}{s} \sum_{i=1}^s \mu^{(i)}$. Also consider any feasible solution $\bar{\mathbf{x}}$ for the original (integer) problem (1) that keeps all 1-entries of $\mathbf{x}_{\diamond}^{(i)}$ (i.e., $\mathbf{x}_{\diamond_j}^{(i)} = 1 \implies \bar{\mathbf{x}}_j^{(i)} = 1$). Then, the additive integrality gap of the original problem can be upper bounded as $\Delta \leq s \cdot \Delta^{(0)}(\bar{\mathbf{x}}^{(0)}) + \sum_{i=1}^s \Delta^{(i)}(\bar{\mathbf{x}}^{(i)})$, where for all $i = 0, \dots, s$*

$$\Delta^{(i)}(\bar{\mathbf{x}}^{(i)}) := \underbrace{\|\mu_{\diamond}^{(i)}\|_1 \cdot \|d\mathbf{1} - A^{(i)}(\mathbf{x}_{\diamond}^{(i)} - \bar{\mathbf{x}}^{(i)})\|_{\infty}}_{\text{constraint slackness}} + \sum_{j: \bar{\mathbf{x}}_j^{(i)} > \mathbf{x}_{\diamond_j}^{(i)}} \underbrace{\left(\langle \mu_{\diamond}^{(i)}, \mathbf{a}_j^{(i)} \rangle - c_j^{(i)}\right)}_{\text{reduced cost}}.$$

The proof of this lemma follows the same line of argument as in [5, 13], and is included in Section 4.2 for completeness. We emphasize that controlling the change in the objective value $\langle \mathbf{c}, \mathbf{x} \rangle$ alone is not

sufficient, since the LP relaxation is solved with a modified right-hand side \mathbf{b}_\diamond . In order to relate the rounded solution back to the original problem with right-hand side \mathbf{b} , we require an upper bound on $\text{OPT}_{\text{LP}}(\mathbf{b})$. This bound is obtained via a standard sensitivity analysis argument based on the optimal dual multiplier μ_\diamond . Taken together, Lemma 4.1 decomposes the integrality gap into two components: one arising from reduced costs and the other from constraint slackness.

Proof strategy of Theorem 4.1. Our proof strategy follows a line similar to that of [13]. In the single-block case, i.e., when $s = 1$, and motivated by Lemma 4.1, the construction of a feasible integral solution $\bar{\mathbf{x}}$ that yields the desired bound on Δ proceeds as follows. We first (conceptually) solve $\text{OPT}_{\text{LP}}(\mathbf{b}_\diamond)$ to obtain an optimal solution \mathbf{x}_\diamond , and then round down all fractional entries of \mathbf{x}_\diamond . While this rounding step guarantees integrality, it inevitably increases the constraint slackness. To compensate for this effect, we subsequently flip a carefully selected subset of originally zero variables with small reduced cost in order to absorb the additional slackness introduced by rounding. A key difficulty of this approach is that the gap formula in Lemma 4.1 depends jointly on $(\mathbf{x}_\diamond, \mu_\diamond)$, whose distributions are complicated and correlated with the random data (A, \mathbf{c}) in (1). Therefore, the existence of a sufficiently large collection of such zero variables is highly non-trivial and is guaranteed via a sophisticated probabilistic argument.

Comparison to [13]. In this work, we extend the result of [13] to multi-block packing integer programs with a two-stage structure. If one ignores the block structure of (1) and applies the argument of [13] directly, one obtains $\Delta \leq O_m\left(\exp(\text{poly}(s))\frac{\log^2 n}{n}\right)$ with high probability, which is significantly weaker than the bound stated in Theorem 4.1. An important observation is that if we fix the optimal dual variable μ_\diamond , then all block variables $\mathbf{x}^{(i)}$ become independent of each other in the Lagrangian $L_{\mathbf{b}_\diamond}^{\text{std}}(\mathbf{x}, \mu)$, since all coupling constraints have been dualized. This phenomenon is also reflected in the gap formula in Lemma 4.1, namely, there are no cross terms between different block variables $\mathbf{x}^{(i)}$. In this case, a natural attempt is to apply the argument of [13] to each block separately, which hopefully would yield $\Delta \leq O_m\left(s \log s \frac{\log^2 n}{n}\right)$, as promised in Theorem 4.1. However, this direct approach encounters a technical difficulty. The analysis in [5, 13] depends critically on the magnitude of the optimal dual solution and on the number of zero entries in the optimal primal solution. Both quantities are sensitive to the block structure of (1) and to the number of scenarios s . In fact, using the arguments in [5, 13], one can show that the optimal dual solution and the number of zero entries of the optimal primal solution behave as in the single-block case **on average**, that is, $\left\|\frac{1}{s} \sum_{i=1}^s \mu_\diamond^{(i)}\right\|_\infty \approx \left\|\text{optimal dual variables when } s = 1\right\|_\infty$. Despite this favorable average behavior, this information alone is insufficient to characterize the full distribution of these quantities. Consequently, it does not suffice to reduce the analysis of the multi-block problem (1) to that of the single-block case. Indeed, a naive blockwise application of the analysis in [13] would still result in an exponential dependence on s , yielding a bound of the form $O_m\left(\exp(\text{poly}(s))\frac{\log^2 n}{n}\right)$. Our main contribution is to show that if we impose an additional lower-bound requirement on the right-hand side \mathbf{b} , namely, $\mathbf{b}_r^{(i)} = \beta_r^{(i)} \cdot 2n, \beta_r^{(i)} \in \left(\frac{1}{4}, \frac{1}{2}\right), \forall i \in [s], r \in [m]$, then the average behavior of the magnitude of the optimal dual solution and the number of zero entries in the optimal primal solution accurately reflects their behavior in each individual block. Under this condition, we can apply the single-block analysis to each block and obtain the desired bound stated in Theorem 4.1.

Let $\mathcal{N}_0^{(i)} := \{j \in [n] : \mathbf{x}_\diamond^{(i)} = 0\}$. Now let *Good* be the event that the following happens:

1. For every block $i \in [s] \cup \{0\}$, $\|\mu_\diamond^{(i)}\|_\infty \leq \frac{1}{2\alpha}$.
2. For every block $i \in [s] \cup \{0\}$, if $\|\mu_\diamond^{(i)}\|_1 \geq \frac{f \log n}{2\alpha n}$, then $|\mathcal{N}_0^{(i)}| \geq \frac{n}{2^{m+2}} \cdot \min\left\{\frac{1}{4}\|\mu_\diamond^{(i)}\|_1, 1\right\}$.

Note that in Item 2 of the event *Good*, we require the cardinality of $\mathcal{N}_0^{(i)}$ to scale proportionally with $\|\mu_\diamond^{(i)}\|_1$, unless $\|\mu_\diamond^{(i)}\|_1$ is already bounded by a small number with order $O(\frac{\log n}{n})$. This requirement, while subtle, is essential for the analysis. In later steps, we will sample columns from $\mathcal{N}_0^{(i)}$ to enforce certain structural properties needed in the proof of Lemma 4.5, and the corresponding sampling procedure depends explicitly on the magnitude of the dual vector $\mu_\diamond^{(i)}$. The condition above guarantees that $\mathcal{N}_0^{(i)}$ contains a sufficiently large collection of indices, ensuring that enough samples can be drawn to carry out the rounding argument successfully.

The following lemma is the main technical result of this section. It shows that, conditioned on the event *Good*, the integrality gap Δ is at most $O_m\left(\frac{sf \log^2 n}{n}\right)$ with high probability.

Lemma 4.2 *Let π be the same constant defined in Theorem 4.1. Under the condition of Theorem 4.1, It follows that*

$$\Pr\left(\Delta > \Omega_m\left(\frac{sf \log^2 n}{n}\right) \mid \text{Good}\right) \leq \frac{(s+1)}{n^2} + (s+1)\pi^f.$$

This is the most technical lemma and the proof is deferred to Section 4.6.

Now we need to show that the event *Good* holds with high probability, i.e., that the complement event *Good*^c happens with tiny probability. Notice that this complement event consists of two union of the events:

$$E1 : \exists i \in [s] \cup \{0\} \text{ such that } \|\mu_\diamond^{(i)}\|_\infty > \frac{1}{2\alpha}$$

$$E2 : \forall i \in [s] \cup \{0\}, \|\mu_\diamond^{(i)}\|_\infty \leq \frac{1}{2\alpha} \text{ but } \exists i \in [s] \cup \{0\} \text{ such that } \|\mu_\diamond^{(i)}\|_1 \geq \frac{f \log n}{2\alpha n} \text{ and } |\mathcal{N}_0^{(i)}| < \frac{n}{2^{m+2}} \cdot \min\left\{\frac{1}{4}\|\mu_\diamond^{(i)}\|_1, 1\right\}.$$

The next lemma bounds the probability of the first event *E1*.

Lemma 4.3 *With probability at least $1 - \frac{1}{n^2}$, under the condition of Theorem 4.1, we have $\|\mu_\diamond^{(i)}\|_\infty \leq \frac{1}{2\alpha}$ for all blocks $i \in [s] \cup \{0\}$.*

The second event *E2* is controlled by the following lemma.

Lemma 4.4 *Under the condition of Theorem 4.1, the probability that $\forall i \in [s] \cup \{0\}, \|\mu_\diamond^{(i)}\|_\infty \leq \frac{1}{2\alpha}$ but $\exists i \in [s] \cup \{0\}$ such that $\|\mu_\diamond^{(i)}\|_1 \geq \frac{f \log n}{2\alpha n}$ and $|\mathcal{N}_0^{(i)}| < \frac{n}{2^{m+2}} \cdot \min\left\{\frac{1}{4}\|\mu_\diamond^{(i)}\|_1, 1\right\}$ is at most $(s+1)/n^2$.*

With these results at hand, the desired bound on the integrality gap directly follows:

PROOF OF THEOREM 4.1: Combining Lemmas 4.3 and 4.4, we obtain that the complement event *Good*^c holds with probability at most $\frac{s+2}{n^2}$. Therefore, further using Lemma 4.2, the probability of a large integrality gap can be upper bounded as

$$\begin{aligned} \Pr\left(\Delta > \Omega_m\left(\frac{sf \log^2 n}{n}\right)\right) &= \Pr\left(\Delta > \Omega_m\left(\frac{sf \log^2 n}{n}\right) \mid \text{Good}\right) \Pr(\text{Good}) + \Pr\left(\Delta > \Omega_m\left(\frac{sf \log^2 n}{n}\right) \mid \text{Good}^c\right) \Pr(\text{Good}^c) \\ &\leq \frac{s+1}{n^2} + (s+1)\pi^f + \frac{s+2}{n^2} = \frac{2s+3}{n^2} + (s+1)\pi^f. \end{aligned}$$

This concludes the proof of the theorem. □

4.2 Proof of Lemma 4.1

Recall that the fact that $(\mathbf{x}_\diamond, \mu_\diamond)$ is a saddle point of $L_{\mathbf{b}_\diamond}^{\text{std}}(\mathbf{x}, \mu)$ means that

$$\text{OPT}_{\text{LP}}(\mathbf{b}_\diamond) = \max_{0 \leq \mathbf{x} \leq 1} \min_{\mu \geq 0} L_{\mathbf{b}_\diamond}^{\text{std}}(\mathbf{x}, \mu) = \max_{0 \leq \mathbf{x} \leq 1} L_{\mathbf{b}_\diamond}^{\text{std}}(\mathbf{x}, \mu_\diamond) = \min_{\mu \geq 0} L_{\mathbf{b}_\diamond}^{\text{std}}(\mathbf{x}_\diamond, \mu),$$

that is, \mathbf{x}_\diamond and μ_\diamond are optimal primal and dual solutions, respectively. Since L^{std} is bilinear and the feasible regions are defined by box constraints, the following complementary slackness conditions hold.

- (cs.i) For all $i \in [s]$, we have $\langle \mu_\diamond^{(i)}, \mathbf{b}_\diamond^{(i)} - A^{(0)} \mathbf{x}_\diamond^{(0)} - A^{(i)} \mathbf{x}_\diamond^{(i)} \rangle = 0$ (if we had “ $<$ ” then $\text{OPT}_{\text{LP}}(\mathbf{b}_\diamond) = \min_{\mu \geq 0} L_{\mathbf{b}_\diamond}^{\text{std}}(\mathbf{x}_\diamond, \mu)$ would be $-\infty$, which is impossible, and if we had “ $>$ ” then μ_\diamond would not be optimal).
- (cs.ii) For every i, j , if $c_j^{(i)} - \langle \mu_\diamond^{(i)}, \mathbf{a}_j^{(i)} \rangle < 0$ then $\mathbf{x}_{\diamond j}^{(i)} = 0$, and if $c_j^{(i)} - \langle \mu_\diamond^{(i)}, \mathbf{a}_j^{(i)} \rangle > 0$ then $\mathbf{x}_{\diamond j}^{(i)} = 1$ (both by optimality of $\mathbf{x}_{\diamond j}^{(i)}$).

Recall that the integrality gap is $\Delta := \text{OPT}_{\text{LP}}(\mathbf{b}) - \text{OPT}(\mathbf{b})$. The LP optimum can be upper bounded using the dual vector μ_\diamond as $\text{OPT}_{\text{LP}}(\mathbf{b}) \leq \max_{0 \leq \mathbf{x} \leq 1} L_{\mathbf{b}}^{\text{std}}(\mathbf{x}, \mu_\diamond)$. Moreover, \mathbf{x}_\diamond remains an optimal maximizer for this problem with right-hand side \mathbf{b} , since (cs.ii) shows that the optimal choice of \mathbf{x} depends only on the reduced costs and is therefore independent of the right-hand side. Hence, $\text{OPT}_{\text{LP}}(\mathbf{b}) \leq L_{\mathbf{b}}^{\text{std}}(\mathbf{x}_\diamond, \mu_\diamond)$. On the other hand, the integer optimum $\text{OPT}(\mathbf{b})$ is lower bounded by the value $s \langle \mathbf{c}^{(0)}, \bar{\mathbf{x}}^{(0)} \rangle + \sum_{i \in [s]} \langle \mathbf{c}^{(i)}, \bar{\mathbf{x}}^{(i)} \rangle$ of the feasible integral solution $\bar{\mathbf{x}}$.

Combining these bounds and adding and subtracting $\sum_{i \in [s]} \langle \mu_\diamond^{(i)}, A^{(0)} \bar{\mathbf{x}}^{(0)} + A^{(i)} \bar{\mathbf{x}}^{(i)} \rangle$, we obtain

$$\begin{aligned} \Delta &\leq s \cdot \langle \mathbf{c}^{(0)} - (\mu_\diamond^{(0)})^\top A^{(0)}, \mathbf{x}_\diamond^{(0)} - \bar{\mathbf{x}}^{(0)} \rangle + \sum_{i \in [s]} \langle \mathbf{c}^{(i)} - (\mu_\diamond^{(i)})^\top A^{(i)}, \mathbf{x}_\diamond^{(i)} - \bar{\mathbf{x}}^{(i)} \rangle \\ &\quad + \sum_{i \in [s]} \langle \mu_\diamond^{(i)}, \mathbf{b}^{(i)} - A^{(0)} \bar{\mathbf{x}}^{(0)} - A^{(i)} \bar{\mathbf{x}}^{(i)} \rangle. \end{aligned} \quad (13)$$

We next bound the first term in (13). Recall that $\mathbf{a}_j^{(0)}$ denotes the j th column of the matrix $A^{(0)}$, we have

$$\begin{aligned} \langle \mathbf{c}^{(0)} - (\mu_\diamond^{(0)})^\top A^{(0)}, \mathbf{x}_\diamond^{(0)} - \bar{\mathbf{x}}^{(0)} \rangle &= \sum_j \left(c_j^{(0)} - \langle \mu_\diamond^{(0)}, \mathbf{a}_j^{(0)} \rangle \right) \cdot (\mathbf{x}_{\diamond j}^{(0)} - \bar{\mathbf{x}}_j^{(0)}) \\ &\leq \sum_{j: \bar{\mathbf{x}}_j^{(0)} > \mathbf{x}_{\diamond j}^{(0)}} \left(\langle \mu_\diamond^{(0)}, \mathbf{a}_j^{(0)} \rangle - c_j^{(0)} \right), \end{aligned} \quad (14)$$

where we justify the contribution of each term j to the last inequality as follows. First, when $\mathbf{x}_{\diamond j}^{(0)} = 1$, by hypothesis we also have $\bar{\mathbf{x}}_j^{(0)} = 1$, so $\mathbf{x}_{\diamond j}^{(0)} - \bar{\mathbf{x}}_j^{(0)} = 0$ and the term j does not contribute to the sum. Now when $\mathbf{x}_{\diamond j}^{(0)} < 1$, (cs.ii) discussed above gives that $c_j^{(0)} - \langle \mu_\diamond^{(0)}, \mathbf{a}_j^{(0)} \rangle \leq 0$, and so swapping the differences the term becomes $(\langle \mu_\diamond^{(0)}, \mathbf{a}_j^{(0)} \rangle - c_j^{(0)}) \cdot (\bar{\mathbf{x}}_j^{(0)} - \mathbf{x}_{\diamond j}^{(0)}) \leq (\langle \mu_\diamond^{(0)}, \mathbf{a}_j^{(0)} \rangle - c_j^{(0)}) \cdot \bar{\mathbf{x}}_j^{(0)}$. When $\bar{\mathbf{x}}_j^{(0)} = 0$, the j th term is then non-positive and can be dropped, and when $\bar{\mathbf{x}}_j^{(0)} = 1$ (which is precisely the case $\bar{\mathbf{x}}_j^{(0)} > \mathbf{x}_{\diamond j}^{(0)}$) the j th term can be upper bounded as $\langle \mu_\diamond^{(0)}, \mathbf{a}_j^{(0)} \rangle - c_j^{(0)}$, as desired.

The same argument allows us to bound the second term in (13) analogously as

$$\langle \mathbf{c}^{(i)} - (\mu_\diamond^{(i)})^\top A^{(i)}, \mathbf{x}_\diamond^{(i)} - \bar{\mathbf{x}}^{(i)} \rangle \leq \sum_{j: \bar{\mathbf{x}}_j^{(i)} > \mathbf{x}_{\diamond j}^{(i)}} \left(\langle \mu_\diamond^{(i)}, \mathbf{a}_j^{(i)} \rangle - c_j^{(i)} \right). \quad (15)$$

To bound the last term of (13), by (cs.i) discussed above we have that that related term, but with respect to \mathbf{x}_\diamond , is actually 0, namely $\langle \mu_\diamond^{(i)}, \mathbf{b}^{(i)} - A^{(0)}\mathbf{x}_\diamond^{(0)} - A^{(i)}\mathbf{x}_\diamond^{(i)} \rangle = 0$. Thus, to upper bound the last term of (13) we subtract from it $\langle \mu_\diamond^{(i)}, \mathbf{b}^{(i)} - A^{(0)}\mathbf{x}_\diamond^{(0)} - A^{(i)}\mathbf{x}_\diamond^{(i)} \rangle$ for all i to obtain that it is at most

$$\begin{aligned} \sum_{i \in [s]} \left\langle \mu_\diamond^{(i)}, \mathbf{b}^{(i)} - A^{(0)}\bar{\mathbf{x}}^{(0)} - A^{(i)}\bar{\mathbf{x}}^{(i)} \right\rangle &= \sum_{i \in [s]} \left\langle \mu_\diamond^{(i)}, \mathbf{b}^{(i)} - \mathbf{b}_\diamond^{(i)} - A^{(0)}(\bar{\mathbf{x}}^{(0)} - \mathbf{x}_\diamond^{(0)}) - A^{(i)}(\bar{\mathbf{x}}^{(i)} - \mathbf{x}_\diamond^{(i)}) \right\rangle \\ &\leq \sum_{i \in [s]} \left\| \mu_\diamond^{(i)} \right\|_1 \cdot \left\| 2d\mathbf{1} - A^{(0)}(\bar{\mathbf{x}}^{(0)} - \mathbf{x}_\diamond^{(0)}) - A^{(i)}(\bar{\mathbf{x}}^{(i)} - \mathbf{x}_\diamond^{(i)}) \right\|_\infty \\ &\leq \sum_{i \in [s]} \left\| \mu_\diamond^{(i)} \right\|_1 \cdot \left\| d\mathbf{1} - A^{(0)}(\bar{\mathbf{x}}^{(0)} - \mathbf{x}_\diamond^{(0)}) \right\|_\infty \\ &\quad + \sum_{i \in [s]} \left\| \mu_\diamond^{(i)} \right\|_1 \cdot \left\| d\mathbf{1} - A^{(i)}(\bar{\mathbf{x}}^{(i)} - \mathbf{x}_\diamond^{(i)}) \right\|_\infty. \end{aligned}$$

Employing the the previous bound and those of (14) and (15) to (13) concludes the proof of Lemma 4.1.

4.3 Bounding the Size of the Optimal Dual (Proof of Lemma 4.3)

To simplify the notation, let $\xi := \frac{2}{4\beta_{\min}-1} = \frac{1}{2\alpha} > 1$. We want to prove that with probability at least $1 - \frac{1}{n^2}$, we have $\|\mu_\diamond^{(i)}\|_\infty \leq \xi$ for all blocks $i \in [s] \cup \{0\}$. We first note that it suffices to consider $i \in [s]$, that is, the case $i = 0$ comes for free: this is because $\mu_\diamond^{(0)}$ is defined as the average $\mu_\diamond^{(0)} = \frac{1}{s} \sum_{i=1}^s \mu_\diamond^{(i)}$ and, since norms are convex, by Jensen's inequality we have $\|\mu_\diamond^{(0)}\|_\infty \leq \frac{1}{s} \sum_{i=1}^s \|\mu_\diamond^{(i)}\|_\infty$.

To prove the desired result, we say that an instance of our problem is **typical** if every row $r \in [m]$ satisfies the following:

1. $\sum_{j \in [n]} A_{r,j}^{(0)} \leq \frac{n}{2} + \sqrt{n \ln snm}$;
2. $\left| \left\{ j \in [n] : \mathbf{c}_j^{(i)} \geq \xi \cdot A_{r,j}^{(i)} \right\} \right| \leq \frac{n}{2\xi} + \sqrt{n \ln snm}$, for all blocks $i \in [s]$.

Notice that, for a fixed row $r \in [m]$, since the entry $A_{r,j}^{(0)}$ is uniformly distributed in $[0, 1]$, in expectation we have $\mathbb{E} \sum_{j \in [n]} A_{r,j}^{(0)} = \frac{n}{2}$; moreover, since $\mathbf{c}_j^{(i)}$ is also uniformly distributed in $[0, 1]$, the probability that $\mathbf{c}_j^{(i)} \geq \xi \cdot A_{r,j}^{(i)}$ can be seen to be $\frac{1}{2\xi}$ (this uses the fact $\xi \geq 1$), and so we expect $\frac{n}{2\xi}$ of the j 's to satisfy this inequality. A quick application of Chernoff's bound show that the extra slack in the definition of a **typical** block is enough to guarantee that this property holds with high probability for all rows and blocks.

Claim 4.2 *If $sm \geq 2$, with probability at least $1 - \frac{1}{n^2}$ the instance is **typical**.*

PROOF: Fix a row $r \in [m]$. Since $A_{r,j}^{(0)}$ is uniformly distributed in $[0, 1]$, we have $\mathbb{E} \sum_{j \in [n]} A_{r,j}^{(0)} = \frac{n}{2}$, and so Chernoff bound (Lemma A.1 in Appendix A) gives

$$\Pr \left(\sum_{j \in [n]} A_{r,j}^{(0)} > \frac{n}{2} + \sqrt{n \ln snm} \right) \leq e^{-2 \ln snm} = \frac{1}{(snm)^2}.$$

Taking a union bound over all rows, Item 1 in the definition of a typical instance holds with probability at least $1 - \frac{1}{(sn)^2 m}$. Also, fix a row $r \in [m]$ and a block $i \in [s]$, and let X_j be the indicator that

$\mathbf{c}_j^{(i)} \geq \xi \cdot A_{r,j}^{(i)}$, so $\sum_{j \in [n]} X_j = |j : \mathbf{c}_j^{(i)} \geq \xi \cdot A_{r,j}^{(i)}|$. Since $(\mathbf{c}_j, A_{r,j}^{(i)})$ is uniformly distributed in the square $[0, 1]^2$ and $\xi \geq 1$, $\Pr(X_i = 1) = \frac{1}{2\xi}$ (i.e., the area of the triangle $[0, 1]^2 \cap \{(y, x) : y \geq \xi x\}$), and again by Chernoff bound $\Pr(\sum_{j \in [n]} X_j \geq \frac{n}{2\xi} + \sqrt{n \ln snm}) \leq \frac{1}{(snm)^2}$. Taking a union bound over all $r \in [m]$ and $i \in [s]$, Item 2 in the definition of a typical instance holds with probability at least $1 - \frac{1}{sn^2m}$. Taking a final union bound over the two items, we conclude the proof of the claim. \square

Then Lemma 4.3 follows from the following bound on the optimal dual for typical instances.

Claim 4.3 *If the instance is **typical**, under the assumption that $2\sqrt{n \ln snm} + 6\alpha m \log n < n \cdot (\beta_{\min} - \frac{1}{4})$, then for all blocks $i \in [s]$ we have $\|\mu_\diamond^{(i)}\|_\infty \leq \xi$.*

PROOF: Let $\text{DUAL}_{\mathbf{b}_\diamond}(\mu) := \max_{0 \leq \mathbf{x} \leq 1} L_{\mathbf{b}_\diamond}^{\text{std}}(\mathbf{x}, \mu)$. The fact that μ_\diamond is a saddle point for $L_{\mathbf{b}_\diamond}^{\text{std}}$ means that it is an argmin of $\text{DUAL}_{\mathbf{b}_\diamond}(\cdot)$.

By means of contradiction, assume that for some block $\bar{i} \in [s]$ we have $\|\mu^{(\bar{i})}\|_\infty > \xi$; without loss of generality, assume this is attained at the first coordinate, namely $\mu_1^{(\bar{i})} > \xi$. First, notice that $\text{DUAL}_{\mathbf{b}_\diamond}(\mu)$ has the the following closed form (we use $(v)^+ := \max\{0, v\}$ to denote the positive part of a value v):

$$\text{DUAL}_{\mathbf{b}_\diamond}(\mu) = \sum_{j=1}^n \left(s \cdot \mathbf{c}^{(0)} - \sum_{i=1}^s (\mu^{(i)})^\top A^{(0)} \right)_j^+ + \sum_{i=1}^s \langle \mu^{(i)}, \mathbf{b}_\diamond^{(i)} \rangle + \sum_{i=1}^s \sum_{j=1}^n \left(\mathbf{c}^{(i)} - (\mu^{(i)})^\top A^{(i)} \right)_j^+, \quad (16)$$

which is obtained by rewriting $L_{\mathbf{b}_\diamond}^{\text{std}}$ (defined in equation (11)) as

$$L_{\mathbf{b}_\diamond}^{\text{std}}(\mathbf{x}, \mu) = \left\langle s \cdot \mathbf{c}^{(0)} - \sum_{i=1}^s (A^{(0)})^\top \mu^{(i)}, \mathbf{x}^{(0)} \right\rangle + \sum_{i=1}^s \langle \mu^{(i)}, \mathbf{b}_\diamond^{(i)} \rangle + \sum_{i=1}^s \langle \mathbf{c}^{(i)} - (A^{(i)})^\top \mu^{(i)}, \mathbf{x}^{(i)} \rangle$$

and taking max over $0 \leq \mathbf{x} \leq 1$.

To reach a contradiction, we now construct a solution μ_* with better value $\text{DUAL}_{\mathbf{b}_\diamond}(\mu_*)$ than μ_\diamond , contradicting its optimality: define μ_* to be the same as μ_\diamond except that $\mu_{*1}^{(\bar{i})} = \mu_{\diamond 1}^{(\bar{i})} - \varepsilon$, for some positive ε .

We claim that $\text{DUAL}_{\mathbf{b}_\diamond}(\mu_*) < \text{DUAL}_{\mathbf{b}_\diamond}(\mu_\diamond)$ for small enough $\varepsilon > 0$. The key observation is that the difference these quantities can be bounded as

$$\text{DUAL}_{\mathbf{b}_\diamond}(\mu_*) - \text{DUAL}_{\mathbf{b}_\diamond}(\mu_\diamond) \leq \sum_{j=1}^n \varepsilon A_{1,j}^{(0)} - \varepsilon \mathbf{b}_{\diamond 1}^{(\bar{i})} + \varepsilon \cdot \left| \left\{ j \in [n] : \left(\mathbf{c}^{(\bar{i})} - (\mu_\diamond^{(\bar{i})})^\top A^{(\bar{i})} \right)_j \geq 0 \right\} \right|. \quad (17)$$

which we now justify: Each of the three terms on the RHS come from the three terms on the RHS of (16), with 1) the first coming from the subadditivity of the positive part function $(\cdot)^+$ and the fact $((\mu_*^{(\bar{i})})^\top A^{(0)})_j = ((\mu_\diamond^{(\bar{i})})^\top A^{(0)})_j - \varepsilon A_{1,j}^{(0)}$; 2) the second term following directly from the definition of μ_* ; 3) the last coming from the fact that for $i \neq \bar{i}$ we do not pick up any difference between $\mu_\diamond^{(i)}$ and $\mu_*^{(i)}$ (they are the same by definition) and for $i = \bar{i}$, for small enough ε because of the positive value function $(\cdot)^+$ we only pick up a difference for j 's such that $(\mathbf{c}^{(\bar{i})} - (\mu_\diamond^{(\bar{i})})^\top A^{(\bar{i})})_j \geq 0$, in which case we pick up a difference of $\varepsilon A_{1,j}^{(\bar{i})} \leq \varepsilon$.

To further upper bound the RHS of (17), we use the assumptions at hand. Since the instance is typical, we directly have $\sum_{j=1}^n \varepsilon A_{1,j}^{(0)} \leq \frac{n}{2} + \sqrt{n \ln snm}$. Also, $-\varepsilon \mathbf{b}_{\diamond 1}^{(\bar{i})} = -\varepsilon \mathbf{b}_1^{(\bar{i})} + \varepsilon 2d = -\varepsilon 2n\beta_1^{(\bar{i})} + \varepsilon 2d \leq$

$-\varepsilon 2n\beta_{\min} + \varepsilon 2d$. Finally, the non-negativity of μ_\diamond and A and the assumption $\mu_{\diamond 1}^{(i)} > \xi$ imply that $(\mathbf{c}^{(i)} - (\mu_\diamond^{(i)})^\top A^{(i)})_j \leq (\mathbf{c}^{(i)} - \mu_{\diamond 1}^{(i)} \cdot A_1^{(i)})_j < (\mathbf{c}^{(i)} - \xi \cdot A_1^{(i)})_j$, and so Item 2 in the definition of a typical instance implies that $\varepsilon \cdot |\{j \in [n] : (\mathbf{c}^{(i)} - (\mu_\diamond^{(i)})^\top A^{(i)})_j \geq 0\}| \leq \varepsilon(\frac{n}{2\xi} + \sqrt{n \ln snm})$.

Plugging these bounds on (17), we get

$$\begin{aligned} \text{DUAL}_{\mathbf{b}_\diamond}(\mu_*) - \text{DUAL}_{\mathbf{b}_\diamond}(\mu_\diamond) &\leq \varepsilon \cdot \left(\frac{n}{2} + \frac{n}{2\xi} - 2n\beta_{\min} + 2\sqrt{n \ln snm} + 2d \right) \\ &= \varepsilon \cdot \left(n \cdot \left(\frac{1}{4} - \beta_{\min} \right) + 2\sqrt{n \ln snm} + 2d \right), \end{aligned} \quad (18)$$

where the last inequality uses the definition $\xi = \frac{2}{4\beta_{\min}-1}$. By the definition of d in (12) and the assumption that $2\sqrt{n \ln snm} + 6\alpha m \log n < n \cdot (\beta_{\min} - \frac{1}{4})$, (18) is strictly negative, which contradicts the optimality of μ_\diamond . This concludes the proof of the claim. \square

4.4 Number of Zeros in the Optimal Primal Solution (Proof of Lemma 4.4)

Recall that $\mathcal{N}_0^{(i)}$ denotes the set of coordinates j for which the optimal primal solution $\mathbf{x}_{\diamond j}^{(i)}$ equals 0. In this section, we prove that the probability of the following event E :

1. $\|\mu_\diamond^{(i)}\|_\infty \leq \frac{1}{2\alpha}$ for all blocks $i \in [s] \cup \{0\}$,
2. but there is a block $i \in [s] \cup \{0\}$ such that $\|\mu_\diamond^{(i)}\|_1 \geq \frac{f \log n}{2\alpha n}$ and $|\mathcal{N}_0^{(i)}| < \frac{n}{2^{m+2}} \cdot \min\{\frac{1}{4}\|\mu_\diamond^{(i)}\|_1, 1\}$

is at most $(s+1)/n^2$.

In order to handle the stochastic dependencies between the set $\mathcal{N}_0^{(i)}$ and the optimal dual vector μ_\diamond , we will consider each possible value of μ_\diamond separately and then take a union bound over a discretized set of these possibilities to obtain the desired result.

The first observation is that if the column $(\mathbf{c}_j^{(i)}, \mathbf{a}_j^{(i)})$ has negative reduced cost with respect to the optimal dual μ_\diamond (i.e., $\mathbf{c}_j^{(i)} - \langle \mathbf{a}_j^{(i)}, \mu_\diamond^{(i)} \rangle < 0$), then since $\mathbf{x}_\diamond^{(i)}$ is a saddle point of $L_{\mathbf{b}_\diamond}^{std}$, we must have $\mathbf{x}_{\diamond j}^{(i)} = 0$, and hence $j \in \mathcal{N}_0^{(i)}$. Thus, to control the size of $\mathcal{N}_0^{(i)}$, it suffices to control the number of columns with negative reduced cost. To this end, for a fixed dual vector $\mu^{(i)}$, define the set of columns whose reduced cost with respect to $\mu^{(i)}$ is less than $-\frac{\log n}{\alpha n}$:

$$\overline{R}_{\mu^{(i)}}^{(i)} := \left\{ j \in [n] : \mathbf{c}_j^{(i)} - \langle \mu^{(i)}, \mathbf{a}_j^{(i)} \rangle < -\frac{\log n}{\alpha n} \right\}.$$

The additional slack introduced by considering such columns, rather than those with merely negative reduced cost, will be crucial in the sequel. Next, we construct a discretization of the possible values of μ_\diamond . Recall that an ε -net Λ (with respect to the norm $\|\cdot\|_1$) of a set $U \subseteq \mathbb{R}^m$ is a subset $\Lambda \subseteq U$ such that for every $u \in U$, there exists $v \in \Lambda$ satisfying $\|u - v\|_1 \leq \varepsilon$. We construct a $\frac{\log n}{\alpha n}$ -net of the set of dual vectors relevant to the event E , namely $\left\{ \mu \in \mathbb{R}_+^m : \|\mu\|_\infty \leq \frac{1}{2\alpha}, \|\mu\|_1 \geq \frac{f \log n}{2\alpha n} \right\}$. Since the argument is standard, we defer the proof to Appendix B.

Claim 4.4 *There is a $\frac{\log n}{\alpha n}$ -net Λ for the set $\{\mu \in \mathbb{R}_+^m : \|\mu\|_\infty \leq \frac{1}{2\alpha}, \|\mu\|_1 \geq \frac{f \log n}{2\alpha n}\}$, with respect to the norm $\|\cdot\|_1$, of size at most $(\frac{mn}{\log n} + 1)^m \leq (2mn)^m$.*

We now formally connect the newly defined sets $\overline{R}_{\mu^{(i)}}^{(i)}$ and Λ with the event of interest E .

Claim 4.5 *If the event E holds, under the assumption that $f \geq 4$, then there is a block $i \in [s] \cup \{0\}$ and a $\mu^{(i)} \in \bar{\Lambda}$ such that $|\bar{R}_{\mu^{(i)}}^{(i)}| < \frac{n}{2^{m+2}} \cdot \min\{\frac{1}{2}\|\mu^{(i)}\|_1, 1\}$.*

PROOF: Suppose the event E holds, and let \bar{i} be a block such that $\|\mu_{\diamond}^{(\bar{i})}\|_1 \geq \frac{f \log n}{2\alpha n}$ and $|\mathcal{N}_0^{(\bar{i})}| < \frac{n}{2^{m+2}} \cdot \min\{\frac{1}{4}\|\mu_{\diamond}^{(\bar{i})}\|_1, 1\}$ (note in addition we have $\|\mu_{\diamond}^{(\bar{i})}\|_{\infty} \leq \frac{1}{2\alpha}$). Let $\mu^{(\bar{i})}$ be a vector in the set Λ such that $\|\mu^{(\bar{i})} - \mu_{\diamond}^{(\bar{i})}\|_1 \leq \frac{\log n}{\alpha n}$. We claim that $\mu^{(\bar{i})}$ satisfies the desired properties, namely $|\bar{R}_{\mu^{(\bar{i})}}^{(\bar{i})}| < \frac{n}{2^{m+2}} \cdot \min\{\frac{1}{2}\|\mu^{(\bar{i})}\|_1, 1\}$.

To see that, we show that $\bar{R}_{\mu^{(\bar{i})}}^{(\bar{i})} \subseteq \mathcal{N}_0^{(\bar{i})}$: Take $j \in \bar{R}_{\mu^{(\bar{i})}}^{(\bar{i})}$; this implies

$$\mathbf{c}_j^{(\bar{i})} + \frac{\log n}{\alpha n} < \langle \mathbf{a}_j^{(\bar{i})}, \mu^{(\bar{i})} \rangle = \langle \mathbf{a}_j^{(\bar{i})}, \mu_{\diamond}^{(\bar{i})} \rangle + \langle \mathbf{a}_j^{(\bar{i})}, \mu^{(\bar{i})} - \mu_{\diamond}^{(\bar{i})} \rangle \leq \langle \mathbf{a}_j^{(\bar{i})}, \mu_{\diamond}^{(\bar{i})} \rangle + \frac{\log n}{\alpha n},$$

the last inequality using the fact $\mathbf{a}_j^{(\bar{i})} \in [0, 1]^m$; reorganizing the terms this implies that $\mathbf{c}_j^{(\bar{i})} - \langle \mathbf{a}_j^{(\bar{i})}, \mu_{\diamond}^{(\bar{i})} \rangle < 0$, and so $j \in \mathcal{N}_0^{(\bar{i})}$, giving the desired result.

Next, we show that $\frac{1}{2}\|\mu_{\diamond}^{(\bar{i})}\|_1 \leq \|\mu^{(\bar{i})}\|_1$: using triangle inequality

$$\|\mu_{\diamond}^{(\bar{i})}\|_1 = \|\mu^{(\bar{i})} + (\mu_{\diamond}^{(\bar{i})} - \mu^{(\bar{i})})\|_1 \leq \|\mu^{(\bar{i})}\|_1 + \|\mu_{\diamond}^{(\bar{i})} - \mu^{(\bar{i})}\|_1 \leq \|\mu^{(\bar{i})}\|_1 + \frac{\log n}{\alpha n} \leq \|\mu^{(\bar{i})}\|_1 + \frac{1}{2}\|\mu_{\diamond}^{(\bar{i})}\|_1,$$

where the last inequality uses the fact $\|\mu_{\diamond}^{(\bar{i})}\|_1 \geq \frac{f \log n}{2\alpha n} \geq \frac{2 \log n}{\alpha n}$; rearranging gives the desired result.

Combining the previous two paragraphs we then get

$$|\bar{R}_{\mu^{(\bar{i})}}^{(\bar{i})}| \leq |\mathcal{N}_0^{(\bar{i})}| \leq \frac{n}{2^{m+2}} \cdot \min\left\{\frac{1}{4}\|\mu_{\diamond}^{(\bar{i})}\|_1, 1\right\} \leq \frac{n}{2^{m+2}} \cdot \min\left\{\frac{1}{2}\|\mu^{(\bar{i})}\|_1, 1\right\},$$

thus proving the claim. \square

We now bound the probability that one of these sets $\bar{R}_{\mu^{(i)}}^{(i)}$ is small, which together with the previous result upper bounds the probability of the event E and gives the proof of Lemma 4.4.

Claim 4.6 *Under the assumption that $f \geq \max\left\{8, \frac{\alpha 2^{m+6} \log(2mn)}{\log n}\right\}$ it holds*

$$\Pr\left(\text{there is } i \in [s] \cup \{0\} \text{ and } \mu^{(i)} \in \Lambda \text{ such that } |\bar{R}_{\mu^{(i)}}^{(i)}| < \frac{n}{2^{m+2}} \cdot \min\left\{\frac{1}{2}\|\mu^{(i)}\|_1, 1\right\}\right) \leq \frac{s+1}{n^2}.$$

PROOF: For each j , notice that whenever $\mathbf{a}_j^{(i)} \in [\frac{1}{2}, 1]^m$ and $\mathbf{c}_j^{(i)} \in [0, \frac{1}{4}\|\mu^{(i)}\|_1)$, we have

$$\mathbf{c}_j^{(i)} - \langle \mathbf{a}_j^{(i)}, \mu^{(i)} \rangle \leq \mathbf{c}_j^{(i)} - \frac{1}{2}\|\mu^{(i)}\|_1 < -\frac{1}{4}\|\mu^{(i)}\|_1 \leq -\frac{\log n}{\alpha n},$$

where the last inequality uses the fact $\|\mu^{(i)}\|_1 \geq \frac{f \log n}{2\alpha n} \geq \frac{4 \log n}{\alpha n}$ since $\mu^{(i)} \in \Lambda$. Thus, when this happens, the index j belongs to the set $\bar{R}_{\mu^{(i)}}^{(i)}$. Since $\mathbf{c}_j^{(i)}$ and $\mathbf{a}_j^{(i)}$ have entries uniformly distributed in $[0, 1]$, this observation implies that j belongs to $\bar{R}_{\mu^{(i)}}^{(i)}$ with probability at least $(\frac{1}{2})^m \cdot \min\{\frac{1}{4}\|\mu^{(i)}\|_1, 1\}$.

Therefore, the expected size of this set, $\mathbb{E}|\bar{R}_{\mu^{(i)}}^{(i)}|$, is at least $n \cdot (\frac{1}{2})^m \cdot \min\{\frac{1}{4}\|\mu^{(i)}\|_1, 1\}$, and further applying the multiplicative Chernoff bound (Lemma A.1) over the indicator that each $j \in \bar{R}_{\mu^{(i)}}^{(i)}$ we get

$$\Pr\left(|\bar{R}_{\mu^{(i)}}^{(i)}| < \frac{n}{2^{m+1}} \cdot \min\left\{\frac{1}{4}\|\mu^{(i)}\|_1, 1\right\}\right) \leq \exp\left(-\frac{n}{2^{m+4}} \cdot \min\left\{\frac{1}{2}\|\mu^{(i)}\|_1, 1\right\}\right) \leq \frac{1}{(2mn)^{m+2}},$$

where the last inequality uses again that $\|\mu^{(i)}\|_1 \geq \frac{f \log n}{2\alpha n} \geq \frac{2^{m+5}(m+2) \log(2mn)}{n}$ since $\mu^{(i)} \in \Lambda$ and the assumption of f .

This implies that

$$\Pr \left(|\overline{R}_{\mu^{(i)}}| < \frac{n}{2^{m+2}} \cdot \min \left\{ \frac{1}{2} \|\mu^{(i)}\|_1, 1 \right\} \right) \leq \Pr \left(|\overline{R}_{\mu^{(i)}}| < \frac{n}{2^{m+1}} \cdot \min \left\{ \frac{1}{4} \|\mu^{(i)}\|_1, 1 \right\} \right) \leq \frac{1}{(2mn)^{m+2}}$$

Finally, taking a union bound over all blocks $i \in [s]$ and all vectors $\mu^{(i)} \in \Lambda$, we obtain

$$\begin{aligned} & \Pr \left(\text{there is } i \in [s] \cup \{0\} \text{ and } \mu^{(i)} \in \Lambda \text{ such that } |\overline{R}_{\mu^{(i)}}| < \frac{n}{2^{m+2}} \cdot \min \left\{ \frac{1}{2} \|\mu^{(i)}\|_1, 1 \right\} \right) \\ & \leq \sum_{i \in [s] \cup \{0\}} \sum_{\mu^{(i)} \in \Lambda} \Pr \left(|\overline{R}_{\mu^{(i)}}| < \frac{n}{2^{m+2}} \cdot \min \left\{ \frac{1}{2} \|\mu^{(i)}\|_1, 1 \right\} \right) \leq (s+1) \cdot |\Lambda| \cdot \frac{1}{(2mn)^{m+2}} \leq \frac{s+1}{n^2}. \end{aligned}$$

This concludes the proof of the claim. \square

4.5 Discrepancy Lemma

We next introduce a discrepancy lemma due to Dyer and Frieze, which is central to the proof of the integrality gap bound. We present a slightly simplified version of the result that is cleaner and sufficient for our purposes. We also note that this discrepancy lemma may be further improved under alternative assumptions on the stochastic model using more sophisticated techniques; for related developments, we refer the reader to [5, 6].

Lemma 4.7 [13] *Fix constants $a, M, \varepsilon > 0$. Then for k being at least a sufficiently large integer (depending on a, α, M) the following holds.*

Let Y^1, \dots, Y^{2k} be $2k$ random vectors in $[-a, a]^m$ where all the coordinates, across all vectors, are independent and identically distributed. Further assume that each coordinate Y_i^j satisfies $\mathbb{E}Y_i^j = 0$ and $\text{Var}(Y_i^j) = 1$ and has a continuous density g satisfying $g(x) \leq M$. For every deterministic vector $\mathbf{v} \in \mathbb{R}^m$ satisfying $\|\mathbf{v}\|_\infty \leq k^{1/2-\varepsilon}$, we have

$$\mathbb{P} \left(\exists K \subseteq \{1, 2, \dots, 2k\} : |K| = k, \left\| \sum_{j \in K} Y^j - \mathbf{v} \right\|_\infty \leq \frac{3k}{4^{k/m}} \right) \geq \frac{1}{2^{m+2}}.$$

4.6 Existence of Good Rounding (Proof of Lemma 4.2)

Recall the definition of the event *Good*: it occurs whenever both of the following conditions hold:

1. For every block $i \in [s] \cup \{0\}$, we have $\|\mu_\diamond^{(i)}\|_\infty \leq \frac{1}{2\alpha}$.
2. For every block $i \in [s] \cup \{0\}$, if $\|\mu_\diamond^{(i)}\|_1 \geq \frac{f \cdot \log n}{2\alpha n}$, then $|\mathcal{N}_0^{(i)}| \geq \frac{n}{2^{m+2}} \cdot \min \left\{ \frac{1}{4} \|\mu_\diamond^{(i)}\|_1, 1 \right\}$.

Our goal is to prove that, conditioned on the event *Good*, with probability at least $1 - \frac{s+1}{n^2} - (s+1)\pi^f$ we obtain a small integrality gap, namely, $\Delta \leq O_m \left(\frac{sf \log^2 n}{n} \right)$. Recall that Lemma 4.1 states that for

any feasible solution $\bar{\mathbf{x}}$ to the original (integer) problem (1) that preserves all 1-entries of $\mathbf{x}_\diamond^{(i)}$ (i.e., if $\mathbf{x}_{\diamond j}^{(i)} = 1$ then $\bar{\mathbf{x}}_j^{(i)} = 1$), the integrality gap Δ can be bounded by

$$\Delta \leq s \cdot \Delta^{(0)}(\bar{\mathbf{x}}^{(0)}) + \sum_{i=1}^s \Delta^{(i)}(\bar{\mathbf{x}}^{(i)}), \quad (19)$$

where

$$\Delta^{(i)}(\bar{\mathbf{x}}^{(i)}) = \|\mu_\diamond^{(i)}\|_1 \cdot \|A^{(i)}(\mathbf{x}_\diamond^{(i)} - \bar{\mathbf{x}}^{(i)}) + d\mathbf{1}\|_\infty + \sum_{j: \bar{\mathbf{x}}_j^{(i)} > \mathbf{x}_{\diamond j}^{(i)}} \left(\langle \mu_\diamond^{(i)}, \mathbf{a}_j^{(i)} \rangle - c_j^{(i)} \right). \quad (20)$$

Thus, it suffices to show that under *Good* there exists a rounding $\bar{\mathbf{x}}$ such that all the quantities $\Delta^{(i)}(\bar{\mathbf{x}}^{(i)})$ are small. To this end, we further refine the conditioning on *Good*. Specifically, we will further condition on (i) the set of coordinates where the optimal primal Lagrangian solution \mathbf{x}_\diamond equals 0, and (ii) the exact data (i.e., $\mathbf{c}_j^{(i)}$ and $A^{(i),j}$) on the remaining coordinates. As we will see later (Lemma 4.6), this conditioning fixes the optimal dual vector μ_\diamond , but leaves the columns on the nonzero coordinates independent and uniformly distributed, subject only to having nonpositive reduced cost. This leaves enough randomness for us to find the desired rounding with high probability by using some of these columns. To formalize this, let $\mathcal{N}_0 := \{(i, j) : \mathbf{x}_{\diamond j}^{(i)} = 0\}$ be the set of coordinates where the optimal primal Lagrangian solution equals 0.

Lemma 4.5 *Consider a conditioning of the set \mathcal{N}_0 and of the column data $((\mathbf{c}_j^{(i)}, \mathbf{a}_j^{(i)}))_{(i,j) \notin \mathcal{N}_0}$ in the complement of \mathcal{N}_0 that belongs to the event *Good*. Under the condition of Theorem 4.1, then with (conditional) probability at least $1 - \frac{(s+1)}{n^2} - (s+1)\pi^f$ there exists an integer solution $\bar{\mathbf{x}} \in \{0, 1\}^{n \cdot (s+1)}$ that:³*

1. Keeps all 1-entries of $\mathbf{x}_\diamond^{(i)}$;
2. Is feasible for (1) ;
3. Has $\Delta^{(i)}(\bar{\mathbf{x}}^{(i)}) \leq O_m\left(\frac{f \log^2 n}{n}\right)$ for all $i \in [s] \cup \{0\}$;

where $\pi := \left(1 - \frac{1}{2^{m+2}}\right)^{\frac{\alpha^m}{2^{m+6m}}} \in (0, 1)$.

This result directly implies the desired bound on the integrality gap, i.e., Lemma 4.2.

PROOF OF LEMMA 4.2: Consider a conditioning of \mathcal{N}_0 and $((\mathbf{c}_j^{(i)}, \mathbf{a}_j^{(i)}))_{(i,j) \notin \mathcal{N}_0}$ that belongs to the event *Good*; by employing the bound from Lemma 4.5 on (19), under this conditioning we have

$$\Delta \leq O_m\left(\frac{sf \log^2 n}{n}\right)$$

with probability at least $1 - \frac{s+1}{n^2} - (s+1)\pi^f$. Taking expectation over all the conditionings of \mathcal{N}_0 and $((\mathbf{c}_j^{(i)}, \mathbf{a}_j^{(i)}))_{(i,j) \notin \mathcal{N}_0}$ that belongs to the event *Good* gives

$$\Pr\left(\Delta \leq O_m\left(\frac{sf \log^2 n}{n}\right) \mid \text{Good}\right) \geq 1 - \frac{s+1}{n^2} - (s+1)\pi^f,$$

as desired. □

So for the remainder of the section we prove Lemma 4.5.

³Actually because of Lemma 4.6, this result does not hold for all the conditionings of the set \mathcal{N}_0 and $((\mathbf{c}_j^{(i)}, \mathbf{a}_j^{(i)}))_{(i,j) \notin \mathcal{N}_0}$, there is a set of them of measure 0 where the result may not hold. Since such set of measure 0 will not affect the use of this lemma, we chose to not include this subtlety in the statement.

4.6.1 Proof of Lemma 4.5

We start by formalizing the previous claim that conditioning on the index set \mathcal{N}_0 and on the columns $((\mathbf{c}_j^{(i)}, A_j^{(i)}))_{(i,j) \notin \mathcal{N}_0}$ at the remaining indices fixes the optimal dual vector μ_\diamond (almost surely), while leaving the columns on the zero coordinates independent and uniformly distributed, subject only to having nonpositive reduced cost. To this end, for a dual vector $\mu^{(i)} \in \mathbb{R}_+^m$, define

$$R_{\mu^{(i)}}^{(i)} := \{(z, \mathbf{y}) \in [0, 1]^{1+m} : z - \langle \mathbf{y}, \mu^{(i)} \rangle \leq 0\}.$$

Thinking of (z, \mathbf{y}) as the data of a column in our problem (where \mathbf{z} represents the gain), $R_{\mu^{(i)}}^{(i)}$ is the set of column values that yield nonpositive reduced cost with respect to the dual vector $\mu^{(i)}$. We will use the following result from [5, 13]. Although the result is stated for a single-block instance, it applies directly here by viewing (1) as a single-block instance.

Lemma 4.6 [5, 13] *Conditional on the value of the set \mathcal{N}_0 , the data $((\mathbf{c}_j^{(i)}, \mathbf{a}_j^{(i)}))_{(i,j) \notin \mathcal{N}_0}$ in the remaining indices uniquely determines the optimal primal/dual Lagrangian solutions $\mathbf{x}_\diamond, \mu_\diamond$ almost surely. If we further condition on an exact value of $((\mathbf{c}_j^{(i)}, \mathbf{a}_j^{(i)}))_{(i,j) \notin \mathcal{N}_0}$ that determines $\mathbf{x}_\diamond, \mu_\diamond$ uniquely, then the columns $(\mathbf{c}_j^{(i)}, \mathbf{a}_j^{(i)})$ for $(i, j) \in \mathcal{N}_0$ are independent and each is uniformly distributed in the set $R_{\mu_\diamond}^{(i)}$.*

For the remainder of the proof, we condition on the value of the index set \mathcal{N}_0 and on the data $((\mathbf{c}_j^{(i)}, \mathbf{a}_j^{(i)}))_{(i,j) \notin \mathcal{N}_0}$ for the remaining columns, such that the pair $(\mathbf{x}_\diamond, \mu_\diamond)$ is uniquely defined and the event *Good* holds. That is, for every block $i \in [s] \cup \{0\}$ we have $\|\mu_\diamond^{(i)}\|_\infty \leq \frac{1}{2\alpha}$, and whenever $\|\mu_\diamond^{(i)}\|_1 \geq \frac{f \log n}{2\alpha n}$ we also have $|\mathcal{N}_0^{(i)}| \geq \frac{n}{2^{m+2}} \cdot \min\left\{\frac{1}{4}\|\mu_\diamond^{(i)}\|_1, 1\right\}$. We denote this conditioning (i.e., the set of outcomes satisfying these conditions) by *Cond*.

We now construct the desired rounded solution $\bar{\mathbf{x}} = (\bar{\mathbf{x}}^{(0)}, \dots, \bar{\mathbf{x}}^{(s)}) \in \{0, 1\}^{(s+1) \cdot n}$ by the following two steps: (i) rounding down the fractional coordinates of \mathbf{x}_\diamond , and (ii) setting some of the originally zero coordinates to 1. More precisely, let $\lfloor \mathbf{x}_\diamond^{(i)} \rfloor$ denote the vector obtained by rounding down $\mathbf{x}_\diamond^{(i)}$ componentwise, and let $\chi_S \in \{0, 1\}^n$ denote the indicator vector of a set $S \subseteq [n]$. We define $\bar{\mathbf{x}}^{(i)} := \lfloor \mathbf{x}_\diamond^{(i)} \rfloor + \chi_{T^{(i)}}$, for some set $T^{(i)} \subseteq \mathcal{N}_0^{(i)}$ of coordinates that are set to 1.

By construction, the vector $\bar{\mathbf{x}}^{(i)}$ preserves all 1-entries of $\mathbf{x}_\diamond^{(i)}$, which establishes Item 1 of Lemma 4.5. We now construct the sets $T^{(i)}$, for each block $i \in [s] \cup \{0\}$, so as to satisfy

$$A^{(i)} \bar{\mathbf{x}}^{(i)} \leq A^{(i)} \mathbf{x}_\diamond^{(i)} + d\mathbf{1}, \quad (21)$$

$$\Delta^{(i)}(\bar{\mathbf{x}}^{(i)}) \leq O_m \left(\frac{f \log^2 n}{n} \right). \quad (22)$$

The second inequality is precisely Item 3 of Lemma 4.5. The first inequality implies Item 2, namely the feasibility of $\bar{\mathbf{x}}$. Indeed, recall that the original right-hand side for scenario i in the packing problem is $\mathbf{b}^{(i)} = \mathbf{b}_\diamond^{(i)} + 2d\mathbf{1}$. Since \mathbf{x}_\diamond is feasible for the modified right-hand side, we have

$$A^{(0)} \mathbf{x}_\diamond^{(0)} + A^{(i)} \mathbf{x}_\diamond^{(i)} \leq \mathbf{b}_\diamond^{(i)}.$$

Therefore, inequality (21) implies

$$A^{(0)} \bar{\mathbf{x}}^{(0)} + A^{(i)} \bar{\mathbf{x}}^{(i)} \leq \mathbf{b}_\diamond^{(i)} + 2d\mathbf{1} = \mathbf{b}^{(i)},$$

which establishes the feasibility of $\bar{\mathbf{x}}$.

Now for each block $i \in [s] \cup \{0\}$, we define the set $T^{(i)}$ to satisfy (21) and (22) (with high probability), breaking into two cases based on the size of $\|\mu_\diamond^{(i)}\|_1$ as follows.

Case 1 if $\|\mu_\diamond^{(i)}\|_1 < \frac{f \log n}{2\alpha n}$: Here we choose the set $T^{(i)}$ to be empty, that is, no additional coordinate is set to 1 in $\bar{\mathbf{x}}^{(i)}$, so that $\bar{\mathbf{x}}^{(i)} = \lfloor \mathbf{x}_\diamond^{(i)} \rfloor$. Clearly, $\bar{\mathbf{x}}^{(i)}$ satisfies (21). Moreover, in this case $\Delta^{(i)}(\bar{\mathbf{x}}^{(i)})$ (recall its definition in (20)) reduces to

$$\begin{aligned} \Delta^{(i)}(\bar{\mathbf{x}}^{(i)}) &= \|\mu_\diamond^{(i)}\|_1 \cdot \|A^{(i)}(\mathbf{x}_\diamond^{(i)} - \lfloor \mathbf{x}_\diamond^{(i)} \rfloor) + d\mathbf{1}\|_\infty \\ &< \frac{f \log n}{2\alpha n} \cdot \|A^{(i)}(\mathbf{x}_\diamond^{(i)} - \lfloor \mathbf{x}_\diamond^{(i)} \rfloor) + d\mathbf{1}\|_\infty, \end{aligned}$$

where the inequality follows from the assumption $\|\mu_\diamond^{(i)}\|_1 < \frac{f \log n}{2\alpha n}$.

To upper bound the remaining term, we first observe that the optimal solution \mathbf{x}_\diamond is a *basic* optimal solution (by uniqueness). By standard arguments, $\mathbf{x}_\diamond^{(i)}$ therefore has at most m fractional coordinates; consequently, $\mathbf{x}_\diamond^{(i)} - \lfloor \mathbf{x}_\diamond^{(i)} \rfloor$ has at most m nonzero entries (see Appendix C for details). Since all entries of $A^{(i)}$ are at most 1, this implies

$$0 \leq A^{(i)}(\mathbf{x}_\diamond^{(i)} - \lfloor \mathbf{x}_\diamond^{(i)} \rfloor) \leq m\mathbf{1}. \quad (23)$$

Substituting (23) into the previous bound yields

$$\Delta^{(i)}(\bar{\mathbf{x}}^{(i)}) < \frac{f \log n}{2\alpha n} \cdot (m + d) \leq \frac{f \log n}{2\alpha n} \cdot 4m \log n \leq O_m\left(\frac{f \log^2 n}{n}\right),$$

where the second inequality uses the definition of d in (12), which implies $d \leq 3\alpha m \log n \leq 3m \log n$. Thus, $\bar{\mathbf{x}}^{(i)}$ satisfies (22), completing the analysis of this case.

Case 2 if $\|\mu_\diamond^{(i)}\|_1 \geq \frac{f \log n}{2\alpha n}$: In this case the construction of the set $T^{(i)}$ is significantly more involved.

From Lemma 4.6, we know that, under the conditioning *Cond*, the columns indexed by $j \in \mathcal{N}_0^{(i)}$ of block i remain independent and uniformly distributed over the set $R_{\mu_\diamond^{(i)}}^{(i)}$ of column values with nonpositive reduced cost with respect to $\mu_\diamond^{(i)}$. However, the nonpositive reduced cost constraint still induces undesirable correlations among the entries of columns in $\mathcal{N}_0^{(i)}$. To obtain additional structure, we therefore restrict attention to a box-like subset of $R_{\mu_\diamond^{(i)}}^{(i)}$ that consists only of column values whose reduced costs are not too negative, so that their contribution to $\Delta^{(i)}$ remains controlled.

Claim 4.8 For every $i \in [s] \cup \{0\}$ and vector $\mu^{(i)} \in \mathbb{R}_+^m$ such that $\frac{f \log n}{2\alpha n} \leq \|\mu^{(i)}\|_1$ and $\|\mu^{(i)}\|_\infty \leq \frac{1}{2\alpha}$, there exists a subset $M_{\mu^{(i)}}^{(i)} \subseteq R_{\mu^{(i)}}^{(i)}$ with following properties:

1. (Small reduced cost wrt $\mu^{(i)}$) Every vector $(z, \mathbf{y}) \in M_{\mu^{(i)}}^{(i)}$ satisfies $z - \langle \mathbf{y}, \mu^{(i)} \rangle \geq -\frac{f \log n}{n}$.
2. (Box-like) If $(Z, Y) \in \mathbb{R} \times \mathbb{R}^m$ is a random variable uniformly distributed in $M_{\mu^{(i)}}^{(i)}$, then Y is uniformly distributed in $[\alpha, 2\alpha]^m$.

3. (Relative volume) $\frac{\text{vol}(M_{\mu^{(i)}}^{(i)})}{\text{vol}(R_{\mu^{(i)}}^{(i)})} \geq \frac{f \log n \cdot \alpha^m}{n \cdot \min\{\|\mu_\diamond^{(i)}\|_1, 1\}}$.

PROOF SKETCH: The region $M_{\mu^{(i)}}^{(i)}$ is defined as

$$M_{\mu^{(i)}}^{(i)} := \bigcup_{y \in [\alpha, 2\alpha]^m} \left(\left[\langle y, \mu^{(i)} \rangle - \frac{f \log n}{n}, \langle y, \mu^{(i)} \rangle \right] \times \{y\} \right).$$

That is, we place the interval $[\langle y, \mu^{(i)} \rangle - \frac{f \log n}{n}, \langle y, \mu^{(i)} \rangle]$ on the first coordinate for each completion $y \in [\alpha, 2\alpha]^m$ of the other coordinates. It is easy to verify that this region satisfies the desired properties, for details see Appendix D. \square

Since, conditioned on $Cond$, each column $(\mathbf{c}_j^{(i)}, \mathbf{a}_j^{(i)})$ with $j \in \mathcal{N}_0^{(i)}$ is uniformly distributed over $R_{\mu_\diamond^{(i)}}^{(i)}$, Item 3 of the previous claim implies that such columns satisfy

$$\Pr \left((\mathbf{c}_j^{(i)}, \mathbf{a}_j^{(i)}) \in M_{\mu_\diamond^{(i)}}^{(i)} \mid Cond \right) \geq \frac{f \log n \cdot \alpha^m}{n \cdot \min\{\|\mu_\diamond^{(i)}\|_1, 1\}}.$$

By the definition of $Cond$, the set $\mathcal{N}_0^{(i)}$ contains at least $\frac{n}{2^{m+2}} \cdot \min\left\{\frac{1}{4}\|\mu_\diamond^{(i)}\|_1, 1\right\}$ elements. Therefore, the expected number of columns in $\mathcal{N}_0^{(i)}$ that fall into $M_{\mu_\diamond^{(i)}}^{(i)}$ is at least

$$\frac{f \log n \cdot \alpha^m}{n \cdot \min\{\|\mu_\diamond^{(i)}\|_1, 1\}} \cdot \frac{n}{2^{m+2}} \cdot \min\left\{\frac{1}{4}\|\mu_\diamond^{(i)}\|_1, 1\right\} \geq \frac{f \log n \cdot \alpha^m}{2^{m+4}}.$$

Moreover, using the conditional independence of these columns, we may apply the Chernoff bound (Lemma A.1) conditioned on $Cond$ to obtain a high-probability version of this statement.

Claim 4.9 Let $\mathcal{M}^{(i)} := \left\{ j \in \mathcal{N}_0^{(i)} : (\mathbf{c}_j^{(i)}, \mathbf{a}_j^{(i)}) \in M_{\mu_\diamond^{(i)}}^{(i)} \right\}$ denote the set of columns of $\mathcal{N}_0^{(i)}$ landing in $M_{\mu_\diamond^{(i)}}^{(i)}$. Then

$$\Pr \left(|\mathcal{M}^{(i)}| \geq \frac{f \log n \cdot \alpha^m}{2^{m+5}} \mid Cond \right) \geq 1 - \exp \left(- \frac{f \log n \cdot \alpha^m}{2^{m+7}} \right) \geq 1 - \frac{1}{n^2},$$

where the last inequality follows from the assumption of $f \geq \frac{2^{2m+8}}{\alpha^m}$.

We now further condition on the set of “nice” columns $\mathcal{M}^{(i)}$ being equal to a fixed set $S \subseteq [n]$ that $|S| \geq \frac{f \log n \cdot \alpha^m}{2^{m+5}}$. That is, we define $Cond_S \subseteq Cond$ as the subset of scenarios in $Cond$ for which $\mathcal{M}^{(i)} = S$. Note that, conditioned on $Cond$, the columns indexed by $\mathcal{N}_0^{(i)}$ are independent and uniformly distributed over $R_{\mu_\diamond^{(i)}}^{(i)}$, and since $M_{\mu_\diamond^{(i)}}^{(i)} \subseteq R_{\mu_\diamond^{(i)}}^{(i)}$, conditioning further on $Cond_S$ implies that the columns with indices in S are independent and uniformly distributed over $M_{\mu_\diamond^{(i)}}^{(i)}$. By Item 2 of Claim 4.8, this implies that for each $j \in S$, the vector $\mathbf{a}_j^{(i)}$ is uniformly distributed over $[\alpha, 2\alpha]^m$.

Fix such a set S with cardinality at least $\frac{f \log n \cdot \alpha^m}{2^{m+5}}$, and partition it into $f' := \frac{f \alpha^m}{2^{m+6} m}$ disjoint subsets $S_1, \dots, S_{f'}$, each of size exactly $2m \log n$. (The definition of f guarantees that S is sufficiently large for such a partition.) We will apply the Discrepancy Lemma (Lemma 4.7) to each subset S_t in order to identify a suitable set $T^{(i)} \subseteq S$, which will serve as the additional coordinates of $\bar{\mathbf{x}}^{(i)}$ that are set to 1. Although this construction succeeds for each S_t only with moderate probability, aggregating over all subsets S_t amplifies the probability of finding an appropriate set of coordinates. Finally, define $\mathbf{z}^{(i)} := A^{(i)}(\mathbf{x}_\diamond^{(i)} - \lfloor \mathbf{x}_\diamond^{(i)} \rfloor) + d\mathbf{1}$.

Claim 4.10 For each set S_t , we have the following:

$$\Pr \left(\text{there is } T_t^{(i)} \subseteq S_t \text{ such that } \sum_{j \in T_t^{(i)}} A^{(i),j} \in \left[\mathbf{z}^{(i)} - \frac{\mathbf{1}}{n}, \mathbf{z}^{(i)} \right] \mid \text{Cond}_S \right) \geq \frac{1}{2^{m+2}}.$$

PROOF: Recall that for each $j \in S_t$, conditioned on Cond_S the vector $A^{(i),j}$ is uniformly distributed in $[\alpha, 2\alpha]^m$, and independent across j 's; thus, if we renormalize it as

$$Y^j := \frac{2\sqrt{3}}{\alpha} \left(A^{(i),j} - \frac{3\alpha}{2} \mathbf{1} \right),$$

we see that each coordinate of this vector has mean 0 and variance 1, and the Y_r^j 's are independent across j and r . Then define $k := \frac{|S_t|}{2} = m \log n$, $\theta := \frac{\alpha}{2\sqrt{3}} \cdot \frac{3k}{4^{k/m}}$, and the vector

$$\mathbf{v} := \frac{2\sqrt{3}}{\alpha} \cdot \left(\mathbf{z}^{(i)} - \left(\theta + \frac{3\alpha k}{2} \right) \mathbf{1} \right).$$

Now we can apply the Discrepancy Lemma (Lemma 4.7) to the random vectors Y^j and the target vector \mathbf{v} to obtain that with probability at least $\frac{1}{2^{m+2}}$ there is a set $T_t^{(i)} \subseteq S$ of size k such that $\left\| \sum_{j \in T_t^{(i)}} Y^j - \mathbf{v} \right\|_\infty \leq \frac{3k}{4^{k/m}}$. Unraveling the definitions, this is equivalent to $\sum_{j \in T_t^{(i)}} A^{(i),j} \in [\mathbf{z}^{(i)} - 2\theta \mathbf{1}, \mathbf{z}^{(i)}]$. To conclude the proof, it suffices to show that $2\theta \leq \frac{1}{n}$:

$$2\theta = \frac{\alpha}{\sqrt{3}} \cdot \frac{3m \log n}{4^{\log n}} \leq \frac{m \log n}{n^2} \leq \frac{1}{n},$$

where the last inequality uses the assumption on the size of n .

Actually, in order to apply this discrepancy lemma, we crucially need to satisfy the requirement $\|\mathbf{v}\|_\infty \leq k^{1/2-\varepsilon}$ for some $\varepsilon > 0$. Recalling that $\mathbf{z}^{(i)} := A^{(i)}(\mathbf{x}_\diamond^{(i)} - \lfloor \mathbf{x}_\diamond^{(i)} \rfloor) + d\mathbf{1}$, the definition of \mathbf{v} , and noticing that we defined $d = \theta + \frac{3\alpha k}{2}$ in (12), we have

$$\|\mathbf{v}\|_\infty \leq \frac{2\sqrt{3}}{\alpha} \cdot \left\| A^{(i)}(\mathbf{x}_\diamond^{(i)} - \lfloor \mathbf{x}_\diamond^{(i)} \rfloor) + \left(d - \theta - \frac{3\alpha k}{2} \right) \mathbf{1} \right\|_\infty = \frac{2\sqrt{3}}{\alpha} \cdot \left\| A^{(i)}(\mathbf{x}_\diamond^{(i)} - \lfloor \mathbf{x}_\diamond^{(i)} \rfloor) \right\|_\infty \leq \frac{2\sqrt{3}m}{\alpha},$$

where the last inequality uses again the at most m entries of $\mathbf{x}_\diamond^{(i)}$ are fractional (more precisely, uses (23)). This is at most $k^{1/4} = (m \log n)^{1/4}$, by the assumption on n , which gives $\|\mathbf{v}\|_\infty \leq k^{1/2-\varepsilon}$ (with $\varepsilon = \frac{1}{4}$) as desired. This concludes the proof of the claim. \square

Since each S_t is independent of each other, aggregating over the subsets S_t boosts the probability of finding an appropriate set of coordinates, yielding the following claim, where the definition of f' gives $(1 - \frac{1}{2^{m+2}})^{f'} = (1 - \frac{1}{2^{m+2}})^{\frac{f \alpha^m}{2^{m+6m}}} = \left((1 - \frac{1}{2^{m+2}})^{\frac{\alpha^m}{2^{m+6m}}} \right)^f = \pi^f$.

Claim 4.11 Conditioned on Cond , with probability at least $1 - \frac{1}{n^2} - \pi^f$ there is a set $T^{(i)}$ of size $m \log n$ such that $\sum_{j \in T^{(i)}} A^{(i),j} \in [\mathbf{z}^{(i)} - \frac{\mathbf{1}}{n}, \mathbf{z}^{(i)}]$ and $\mathbf{c}_j^{(i)} - \langle \mathbf{a}_j^{(i)}, \mu_\diamond^{(i)} \rangle \geq -\frac{f \log n}{n}$ for all $j \in T^{(i)}$.

Therefore, we use the set $T^{(i)} \subseteq S$ prescribed by this claim. When its guarantees hold, we immediately obtain (21). Indeed, the binary solution

$$\bar{\mathbf{x}}^{(i)} = \lfloor \mathbf{x}_\diamond^{(i)} \rfloor + \chi_{T^{(i)}}$$

satisfies

$$A^{(i)}\bar{\mathbf{x}}^{(i)} = \sum_{j \in T^{(i)}} A^{(i),j} + A^{(i)}\lfloor \mathbf{x}_\diamond^{(i)} \rfloor \leq \mathbf{z}^{(i)} + A^{(i)}\lfloor \mathbf{x}_\diamond^{(i)} \rfloor = A^{(i)}\mathbf{x}_\diamond^{(i)} + d\mathbf{1}.$$

In addition, this solution satisfies (22):

$$\begin{aligned} \Delta^{(i)}(\bar{\mathbf{x}}^{(i)}) &= \|\mu_\diamond^{(i)}\|_1 \cdot \|A^{(i)}(\mathbf{x}_\diamond^{(i)} - \bar{\mathbf{x}}^{(i)}) + d\mathbf{1}\|_\infty + \sum_{j \in T^{(i)}} \left(\langle \mathbf{a}_j^{(i)}, \mu_\diamond^{(i)} \rangle - \mathbf{c}_j^{(i)} \right) \\ &\leq \frac{m}{2\alpha} \cdot \frac{1}{n} + m \log n \cdot \frac{f \log n}{n} \leq O_m \left(\frac{f \log^2 n}{n} \right), \end{aligned}$$

where we used: (i) $\|\mu_\diamond^{(i)}\|_1 \leq m\|\mu_\diamond^{(i)}\|_\infty \leq \frac{m}{2\alpha}$, with the last inequality following from the definition of the event *Cond*; (ii) the bound from Claim 4.11;

Wrapping up. Combining the two cases $\|\mu_\diamond^{(i)}\|_1 < \frac{f \log n}{2\alpha n}$ and $\|\mu_\diamond^{(i)}\|_1 \geq \frac{f \log n}{2\alpha n}$ considered above, we conclude that for a fixed i , conditioned on *Cond*, with probability at least $1 - \frac{1}{n^2} - \pi^f$, the constructed solution $\bar{\mathbf{x}}^{(i)}$ satisfies both (21) and (22).

Finally, applying a union bound over all $i \in [s] \cup \{0\}$, we obtain that, with probability at least $1 - \frac{s+1}{n^2} - (s+1)\pi^f$, conditioned on *Cond*, we can simultaneously construct solutions $\bar{\mathbf{x}}^{(i)}$ for all blocks i satisfying these properties. This completes the proof of Lemma 4.5.

5 Conclusion

This paper provides a first rigorous average-case analysis of decomposition-based methods for two-stage stochastic integer programs, with a particular focus on dual decomposition and the Branch-and-Price framework. Under a natural stochastic-input model, we show that the Branch-and-Price search tree has quasi-polynomial size with high probability, substantially improving upon the classical worst-case exponential bounds. A key ingredient in the analysis is an average-case bound on the integrality gap of the natural LP relaxation, which shows that the gap is small even when the number of scenarios grows. An important future direction is to develop a comparable average-case analysis for Benders-type methods. While this paper focuses on Branch-and-Price and related dual decomposition schemes, Benders decomposition is another widely used approach for two-stage stochastic integer programs. Establishing average-case guarantees for Benders-type algorithms would be an important step toward a more unified theoretical understanding of why decomposition methods perform so well in practice.

Appendix

A Probabilistic Inequalities

We will need both the additive and multiplicative Chernoff inequalities.

Lemma A.1 [22] *Let X_1, \dots, X_k be independent random variables in $[0, 1]$. Let $S = \sum_i X_i$ be their sum and $\mu := \mathbb{E}S$ its expected value. Then for any $\lambda > 0$ and $\delta \in [0, 1]$ we have:*

$$\Pr(S \geq \mu + \lambda\sqrt{k}) \leq e^{-2\lambda^2}, \quad \Pr(S \leq \mu - \lambda\sqrt{k}) \leq e^{-2\lambda^2}, \quad \text{and} \quad \Pr(S \leq (1 - \delta)\mu) \leq e^{-\frac{\delta^2\mu}{2}}$$

B Proof of Claim 4.4

We construct the desired $\frac{\log n}{\alpha n}$ -net (w.r.t. the norm $\|\cdot\|_1$) of the set $U := \{\mu \in \mathbb{R}_+^m : \|\mu\|_\infty \leq \frac{1}{2\alpha}, \|\mu\|_1 \geq \frac{f \log n}{2\alpha n}\}$ as follows.

First, consider the grid with spacing $\frac{\log n}{\alpha n m}$ of the interval $[0, \frac{1}{2\alpha}]$, plus the endpoint $\frac{1}{2\alpha}$, namely the set $G := ((\frac{\log n}{\alpha n m} \cdot \mathbb{Z}) \cap [0, \frac{1}{2\alpha}]) \cup \{\frac{1}{2\alpha}\} = \{0, \frac{\log n}{\alpha n m}, \frac{2 \log n}{\alpha n m}, \dots, \frac{1}{2\alpha}\}$. Then define the $\frac{\log n}{\alpha n}$ -net as $\Lambda := (G \times G \times \dots \times G) \cap U$, namely taking the m -dimensional product grid and intersecting with the set U .

Clearly Λ is contained in U and satisfies the desired size upper bound $|\Lambda| \leq (\frac{mn}{\log n} + 1)^m$. Now we prove its approximation property, namely consider any $u \in U$ and we will exhibit some $v \in \Lambda$ such that $\|u - v\|_1 \leq \frac{\log n}{\alpha n}$. Indeed, let v be so that its coordinate v_j is the first point in G greater than or equal to u_j ; since $u_j \leq \frac{1}{2\alpha}$ and we include $\frac{1}{2\alpha}$ in the grid G , such point always exists. Moreover, since $v \geq u \geq 0$ we have $\|v\|_1 \geq \|u\|_1 \geq \frac{f \log n}{2\alpha n}$, and so v belongs to Λ . Finally, by construction of the grid we have $u_j \leq v_j \leq u_j + \frac{\log n}{\alpha n m}$, and hence $\|u - v\|_1 \leq \frac{\log n}{\alpha n}$. So Λ is the desired net.

C Number of Fractional Coordinates of $\mathbf{x}_\delta^{(i)}$

We observe the following result:

Lemma C.1 $\mathbf{x}_\delta^{(i)}$ has at most m fractional entries for $i \in [s] \cup \{0\}$.

PROOF: Since $A^{(i)} \in \mathbb{R}^{m \times n}$, for $i \in [s] \cup \{0\}$, there are at most m independent rows of $A^{(i)}$. Every vertex of the LP relaxation is a solution of a linear system where some constraints of (10b) and some constraints of (10c) are active. Since (10c) are bound constraints, each active constraint of (10c) fixes some variable to be 1 or 0. For any such linear system defining a vertex, after eliminating all active constraints of (10c) and the corresponding variables, the remaining linear system takes form of

$$\begin{bmatrix} W^{(01)} & W^{(1)} & 0 & \dots & 0 \\ W^{(02)} & 0 & W^{(2)} & \dots & 0 \\ \dots & 0 & 0 & \dots & 0 \\ W^{(0s)} & 0 & 0 & \dots & W^{(s)} \end{bmatrix} \mathbf{x}_{\text{rem}} = \mathbf{b}_{\text{rem}}$$

where $\mathbf{x}_{\text{rem}}, \mathbf{b}_{\text{rem}}$ are the remaining variables/right-hand-sides and each row of $W^{(i)}$ is from rows of $A^{(i)}$ and each row of $W^{(0i)}$ is from rows of $A^{(0)}$ that are active. Note that there are at most m linearly

independent rows from each $A^{(i)}$ or $A^{(0)}$. Therefore $\text{rank}(W^{(i)}), \text{rank}\left(\begin{bmatrix} W^{(01)} \\ W^{(02)} \\ \dots \\ W^{(0s)} \end{bmatrix}\right) \leq m$. Since the above system is invertible, each $W^{(i)}$ and $W^{(0i)}$ has at most m columns. Therefore at most m entries of $\mathbf{x}_\diamond^{(i)}$ are included in \mathbf{x}_{rem} , which is potentially set to be fractional. \square

D Proof of Claim 4.8

We first verify that indeed $M_{\mu^{(i)}}^{(i)} \subseteq R_{\mu^{(i)}}^{(i)}$, for which it suffices to check that $M^{(i)}$ is contained in the unit cube $[0, 1]^{1+m}$. This is clearly the case, by definition, for the last m coordinates, so it suffices to verify it for the first coordinate, namely that for every $y \in [\alpha, 2\alpha]^m$, the interval $[\langle y, \mu^{(i)} \rangle - \frac{f \log n}{n}, \langle y, \mu^{(i)} \rangle]$ is contained in $[0, 1]$, which is indeed the case: by the lower bound on $\mu^{(i)}$ we have $\langle y, \mu^{(i)} \rangle - \frac{f \log n}{n} \geq \alpha \|\mu^{(i)}\|_1 - \frac{f \log n}{n} \geq 0$, and by the upper bound on $\mu^{(i)}$ we have $\langle y, \mu^{(i)} \rangle \leq 2\alpha \|\mu^{(i)}\|_\infty \leq 1$.

We now prove the different items in the claim.

Item 1. The bound on the reduced cost follows directly by the definition of the region $M_{\mu^{(i)}}^{(i)}$: if (z, y) belongs to this region, then $z \geq \langle y, \mu^{(i)} \rangle - \frac{f \log n}{n}$, or equivalently, $z - \langle y, \mu^{(i)} \rangle \geq -\frac{f \log n}{n}$, as desired.

Item 2. This also follows directly from the definition of $M_{\mu^{(i)}}^{(i)}$.

Item 3. The volume of $R_{\mu^{(i)}}^{(i)}$ is at most $\min\{\|\mu^{(i)}\|_1, 1\}$, because $(z, y) \in R_{\mu^{(i)}}^{(i)}$ means

$$z - \langle y, \mu^{(i)} \rangle < 0 \implies z < \|y\|_\infty \|\mu^{(i)}\|_1 \leq \|\mu^{(i)}\|_1.$$

and also, by definition, $z \leq 1$. Combining this observation with Item 2 gives Item 3, and concludes the proof of the claim.

References

- [1] Gustavo Angulo, Shabbir Ahmed, and Santanu S Dey. Improving the integer l-shaped method. *INFORMS Journal on Computing*, 28(3):483–499, 2016.
- [2] Akul Bansal and Simge Küçükyavuz. A computational study of cutting-plane methods for multi-stage stochastic integer programs. *arXiv preprint arXiv:2405.02533*, 2024.
- [3] Cynthia Barnhart, Ellis L Johnson, George L Nemhauser, Martin WP Savelsbergh, and Pamela H Vance. Branch-and-price: Column generation for solving huge integer programs. *Operations research*, 46(3):316–329, 1998.
- [4] Natashia Boland, Jeffrey Christiansen, Brian Dandurand, Andrew Eberhard, Jeff Linderoth, James Luedtke, and Fabricio Oliveira. Combining progressive hedging with a frank–wolfe method to compute lagrangian dual bounds in stochastic mixed-integer programming. *SIAM Journal on Optimization*, 28(2):1312–1336, 2018.
- [5] Sander Borst, Daniel Dadush, Sophie Huiberts, and Samarth Tiwari. On the integrality gap of binary integer programs with gaussian data. *Mathematical Programming*, 197(2):1221–1263, 2023.
- [6] Sander Borst, Daniel Dadush, and Dan Mikulincer. Integrality gaps for random integer programs via discrepancy. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1692–1733. SIAM, 2023.

- [7] Mikhail A Bragin. Survey on lagrangian relaxation for milp: importance, challenges, historical review, recent advancements, and opportunities. *Annals of Operations Research*, 333(1):29–45, 2024.
- [8] Rui Chen and James Luedtke. On generating lagrangian cuts for two-stage stochastic integer programs. *INFORMS Journal on Computing*, 34(4):2332–2349, 2022.
- [9] Diego Cifuentes, Santanu S Dey, and Jingye Xu. Lagrangian dual for integer optimization with zero duality gap that admits decomposition. In *International Conference on Integer Programming and Combinatorial Optimization*, pages 184–198. Springer, 2025.
- [10] Guy Desaulniers, Jacques Desrosiers, and Marius M Solomon. *Column generation*, volume 5. Springer Science & Business Media, 2006.
- [11] Santanu S Dey, Yatharth Dubey, and Marco Molinaro. Branch-and-bound solves random binary ips in polytime. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 579–591. SIAM, 2021.
- [12] Santanu S Dey, Marco Molinaro, and Qianyi Wang. Analysis of sparse cutting planes for sparse milps with applications to stochastic milps. *Mathematics of Operations Research*, 43(1):304–332, 2018.
- [13] Martin E Dyer and Alan M Frieze. Probabilistic analysis of the multidimensional knapsack problem. *Mathematics of Operations Research*, 14(1):162–176, 1989.
- [14] Kibaek Kim and Brian Dandurand. Scalable branching on dual decomposition of stochastic mixed-integer programming problems. *Mathematical Programming Computation*, 14(1):1–41, 2022.
- [15] Simge Küçükyavuz and Suvrajeet Sen. An introduction to two-stage stochastic mixed-integer programming. In *Leading Developments from INFORMS Communities*, pages 1–27. INFORMS, 2017.
- [16] Xiao Liu, Simge Küçükyavuz, and James Luedtke. Decomposition algorithms for two-stage chance-constrained programs. *Mathematical Programming*, 157(1):219–243, 2016.
- [17] Marco E Lübbecke. Column generation. *Wiley encyclopedia of operations research and management science*, 17:18–19, 2010.
- [18] Ragheb Rahmaniani, Teodor Gabriel Crainic, Michel Gendreau, and Walter Rei. Theenders decomposition algorithm: A literature review. *European Journal of Operational Research*, 259(3):801–817, 2017.
- [19] Ward Romeijndersa, Yihang Zhangb, and Suvrajeet Senb. Stochastic mixed-integer programming: A survey.
- [20] Nikolaos V Sahinidis. Optimization under uncertainty: state-of-the-art and opportunities. *Computers & chemical engineering*, 28(6-7):971–983, 2004.
- [21] Kaizhao Sun, Mou Sun, and Wotao Yin. Decomposition methods for global solution of mixed-integer linear programs. *SIAM Journal on Optimization*, 34(2):1206–1235, 2024.
- [22] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.
- [23] Jikai Zou, Shabbir Ahmed, and Xu Andy Sun. Stochastic dual dynamic integer programming. *Mathematical Programming*, 175(1):461–502, 2019.