Mathematical Programming manuscript No.
(will be inserted by the editor)

Heiko Röglin · Berthold Vöcking

Smoothed Analysis of Integer Programming

Received: date / Revised version: date

Abstract. We present a probabilistic analysis of integer linear programs (ILPs). More specifically, we study ILPs in a so-called smoothed analysis in which it is assumed that first an adversary specifies the coefficients of an integer program and then (some of) these coefficients are randomly perturbed, e.g., using a Gaussian or a uniform distribution with small standard deviation. In this probabilistic model, we investigate structural properties of ILPs and apply them to the analysis of algorithms. For example, we prove a lower bound on the slack of the optimal solution. As a result of our analysis, we are able to specify the smoothed complexity of classes of ILPs in terms of their worst case complexity. For example, we obtain polynomial smoothed complexity for packing and covering problems with any fixed number of constraints. Previous results of this kind were restricted to the case of binary programs. (A preliminary version appeared in the Proceedings of the 11th IPCO (2005) [13].)

1. Introduction

Many algorithmic problems are hard with respect to worst-case instances but there are algorithms for these problem that work quite efficiently on "typical instances", that is, on instances occurring frequently in practice. Finding an adequate theoretical model for typical instances, however, is a challenging task. A reasonable approach seems to be to represent typical instances in form of a probability distribution on the set of possible inputs. A classical average-case analysis begins with the specification of the input dis-

Department of Computer Science – RWTH Aachen

e-mail: {roeglin,voecking}@cs.rwth-aachen.de

Mathematics Subject Classification (1991): 20E28, 20G40, 20C20

tribution. Usually, this is just a simple uniform distribution. The dilemma with such an approach is that any fixed input distribution can be argued to be not the right, typical one. During the last years there has been an increased interest in more general input models and more robust kinds of probabilistic analyses that do not only hold for particular input distributions. An example for such a concept is the so-called *smoothed analysis* of the Simplex algorithm by Spielman and Teng [14]. They assume that first an adversary specifies the input numbers of an LP and then these adversarial numbers are slightly perturbed at random using a Gaussian distribution with specified standard deviation. Spielman and Teng show that the expected running time of the Simplex algorithm under such random perturbations is polynomially bounded in the size of the input and the reciprocal of the standard deviation. Intuitively, this means that the running time function of the Simplex algorithm shows superpolynomial behavior only at some isolated peaks.

Beier and Vöcking [5] generalize smoothed analysis towards discrete optimization problems. In particular, they study optimization problems that can be represented in form of binary programs. A *linear binary optimization problem* is defined by a set of linear constraints and a linear objective function over some subset $\{0,1\}^n$. By parametrizing which constraints are of stochastic and which are of adversarial nature, it is possible to randomize some of the constraints without destroying the combinatorial structure described by other constraints. Their analysis covers various probability distributions for the choice of the stochastic numbers and includes smoothed analysis with Gaussian and other kinds of perturbation models as a special case. It is shown that a binary optimization problem has polynomial smoothed complexity if and only if it has random pseudopolynomial complexity, i.e., the unary variant of the problem is in ZPP. Other results on the smoothed and average-case analysis of discrete optimization problems can be found, e.g., in [1,4,2,3,6,8,10–12]. All these results are restricted to problems that can be written in form of a binary optimization problem.

In this paper, we extend the results of Beier and Vöcking [5] from binary towards integer linear programs (ILPs). We investigate structural properties of ILPs and, as a result of our analysis, we are able to describe the smoothed complexity of classes of ILPs in terms of their worst case complexity similar to the characterization for binary optimization problems. For example, our characterization implies polynomial smoothed (average) complexity for packing and covering problems with any fixed number of constraints since these classes of ILPs admit pseudopolynomial time algorithms. On the other hand, packing and covering problems with an unbounded number of constraints do not have polynomial smoothed complexity, unless ZPP $=$ NP, as these classes are strongly NP-hard[1].

Outline. In the next section, we define the considered probabilistic model and state our results in a formal way. The probabilistic analysis is presented in Section 2. It is centered around structural properties of integer linear programs, called *winner*, *loser* and *feasibility gap*. In Section 3, we show how to exploit these gaps algorithmically in form of an adaptive rounding scheme increasing the accuracy of calculation until the optimal solution is found. Essentially, Sections 2 and 3 show that the existence of a pseudopolynomial algorithm for a class of ILPs implies that this class has polynomially

[1] An NP-hard problem is called *strongly* NP-*hard* if it remains NP-hard even if all input numbers are encoded in unary (see e.g. [9]).

smoothed complexity. In Section 4, we show that the inverse holds as well, that is, polynomial smoothed complexity implies the existence of a pseudopolynomial algorithm. In our probabilistic analysis, we assume that each variable can take only a polynomial number of different values. In Section 5, we show that this assumption is fulfilled for packing and covering ILPs, with high probability, so that we get a tight characterization of polynomial smoothed complexity for packing and covering ILPs. Finally, extensions of our results and conclusions are given in the Sections 6 and 7.

1.1. Problems and model

Our analysis deals with *integer linear programs (ILPs)*. Without loss of generality we consider maximization programs with $\leq$-constraints of the following standard form:

$$\begin{aligned} \max \quad & c^T x \\ \text{s.t.} \quad & Ax \leq b \\ & x \in \mathcal{D}^n, \end{aligned}$$

where $A \in \mathbb{R}^{k \times n}$, $b \in \mathbb{R}^k$, $c \in \mathbb{R}^n$, and $\mathcal{D} \subset \mathbb{Z}$. In our analysis we consider "classes of ILPs", that is, we place certain restrictions on ILPs. Packing and covering ILPs are good examples for such restrictions. In a *packing ILP* all coefficients are non-negative, the objective is $\max c^T x$ and all constraints are of the form $Ax \leq b$. In a *covering ILP* all coefficients are non-negative as well, the objective is $\min c^T x$ and all constraints are of the form $Ax \geq b$. Both in packing and in covering ILPs, the variables are assumed non-negative, that is, $\mathcal{D} = \mathbb{N}_0$. As another example, one can also place restrictions on the number of allowed constraints. Optimization problems with such restrictions are, e.g., specified in the compendium of NP optimization problems [7]. For example, packing ILPs with only one constraint correspond to the INTEGER KNAPSACK PROBLEM, and packing ILPs with a constant number k of constraints correspond to the MAXIMUM INTEGER k-DIMENSIONAL KNAPSACK PROBLEM.

In our probabilistic analysis, we will assume that some of the constraints and/or the objective function are randomly perturbed. The other components of the ILP are not touched by the randomization. When talking about a *class Π of ILPs*, we mean that Π specifies which restrictions are placed on the ILPs and which components are randomly perturbed.

Description of the probabilistic input model. Smoothed analysis assumes a semi-random input model: First, an adversary specifies all input numbers (coefficients in A and c as well as all thresholds in b), then some of the coefficients and thresholds are randomly perturbed. We assume that all numbers specified by the adversary are from the interval $[-1, 1]$. Observe that this is not a restriction as every ILP can be brought into this form by scaling the linear expressions that violate this assumption. In the considered probabilistic input model one can choose which part of the input is perturbed and which part is left untouched by the randomization. Basically, there are three different choices: either only the coefficients in the objective function or only the coefficients and thresholds in the constraints or both are perturbed. We assume that the selected numbers (which we will call *stochastic numbers/coefficients* in the following) are randomly perturbed by adding an independent random variable to each of them. More general perturbation models to which our analysis can also be applied are discussed in Section 6.

Spielman and Teng use Gaussian perturbations [14]. Following [5], we use a more general perturbation model: The random numbers that are added to the adversarial numbers are drawn according to a specified family of probability distributions satisfying the following conditions. Let $f : \mathbb{R} \to \mathbb{R}_{\geq 0}$ be a density function such that $\sup_s(f(s)) = 1$ and $E = \int_{\mathbb{R}} |s| f(s) ds$ is finite. In words, the random variable described by f has "maximum density equal to 1" and a "finite expected absolute mean value". The function f is called the *perturbation model*. For $\phi \geq 1$, we define f_ϕ by scaling f, that is, $f_\phi(s) = \phi f(s\phi)$, for every $s \in \mathbb{R}$. This way it holds $\sup_{s\in\mathbb{R}}(f_\phi(s)) = \phi$ and $\int_{\mathbb{R}} |s| f_\phi(s) ds = E/\phi$. Now we obtain ϕ-perturbations according to perturbation model f by adding an independent random variable with density f_ϕ to each stochastic coefficient.

For example, one obtains the Gaussian perturbation model from [14] by choosing f to be the Gaussian density with standard deviation $(2\pi)^{-1/2}$. A non-negative feasible region for the random numbers can be obtained, e.g., by choosing f to be the density of the uniform distribution over $[0, 1]$. In [14], the running time is described in terms of the input size and the standard deviation σ. Following [5], we describe the running time in terms of the input size and the density parameter ϕ. For the Gaussian and the uniform distribution these two parameters are closely related: in both cases, ϕ is proportional to $1/\sigma$. Intuitively, ϕ can be seen as a measure specifying how close the probabilistic analysis is to a worst-case analysis. A worst-case instance can be interpreted as a stochastic instance in which the probability mass for each stochastic number is mapped to a single point. Thus, the larger ϕ, the closer we are to a worst-case analysis.

Definition of smoothed complexity. The smoothed complexity of a class Π of ILPs with an associated perturbation model f is given in terms of the input length N and the parameter ϕ. First of all, the definition of the input length needs some clarification as some of the input numbers are assumed to be random variables following continuous probability distributions. These numbers are irrational with probability 1 but we define that each of these numbers has a virtual length of one. (This way, we ensure $N \geq nk$.) The bits of the stochastic numbers can be accessed by asking an oracle in time $O(1)$ per bit. The bits after the binary point of each stochastic number are revealed one by one from left to right. As one of the results of our probabilistic analysis, we will see that $O(\log n)$ revealed bits per number are sufficient to determine the optimal solution with high probability. The deterministic part of the input does not contain irrational numbers and can be encoded in an arbitrary fashion. Let I_N denote the set of possible adversarial inputs for Π of length N. For an instance $I \in I_N$, let $I + f_\phi$ denote the random instance that is obtained by a ϕ-perturbation of I. We say that Π has *polynomial smoothed complexity* under f if and only if it admits a polynomial P and an algorithm A whose running time T satisfies

$$\mathbf{Pr}\left[T(I + f_\phi) \geq P\left(N, \phi, \frac{1}{\varepsilon}\right)\right] \leq \varepsilon \ , \tag{1}$$

for every $N \in \mathbb{N}$, $\phi \geq 1$, $\varepsilon \in (0, 1]$, $I \in \mathcal{I}_N$, that is, with probability at least $1 - \varepsilon$ the running time of $\mathcal{A}$ is polynomially bounded in the input length N, the perturbation parameter ϕ, and the reciprocal of ε. An equivalent way of defining polynomial smoothed complexity is to require the existence of an algorithm $\mathcal{A}$ whose running time T satisfies

$$\exists \alpha, \beta > 0 : \forall \phi \geq 1 : \forall N \in \mathbb{N} : \ \max_{I \in \mathcal{I}_N} \mathbf{E}\left[(T(I + f_\phi))^\alpha\right] \ \leq \ \beta \, \phi N \ .$$

This definition of polynomial smoothed complexity follows more or less the way how polynomial complexity is defined in average-case complexity theory, adding the requirement that the running time should be polynomially bounded not only in N but also in ϕ. Observe that this does not imply that the expected running time is polynomially bounded. To enforce expected polynomial running time, the exponent α in the definition of polynomial smoothed complexity should have been placed outside instead of inside the expectation. The reason for not defining polynomial smoothed complexity based on the expected running time is that this is not a sufficiently robust notion. For example, an algorithm with expected polynomial running time on one machine model might have expected exponential running time on another machine model. In contrast, the above definition yields a notion of polynomial smoothed complexity that does not vary among classes of machines admitting polynomial time simulations among each other. The drawback of this definition is, however, that polynomial smoothed complexity does not imply polynomial expected running time.

1.2. Our results

We show that the smoothed complexity of ILPs can be characterized in terms of their worst-case complexity. For a class of ILPs Π, let Π_u denote the corresponding optimization problem in which the stochastic numbers are assumed to be integers in unary representation instead of randomly chosen real-valued numbers.

Theorem 1. *A class Π of ILPs where each variable can take only a polynomial number of different values, that is $|\mathcal{D}|$ is polynomially bounded in the number of variables, has polynomial smoothed complexity if and only if $\Pi_u \in$ ZPP.*

In other words, Π has polynomial smoothed complexity if it admits a (possibly randomized) algorithm with (expected) pseudopolynomial worst-case running time. If we apply this theorem to ILPs containing at least one perturbed packing constraint then we can even drop the restriction on the set $\mathcal{D}$ as perturbed instances of these problems fulfill this restriction with high probability.

Theorem 2. *A class Π of ILPs with at least one perturbed packing constraint has polynomial smoothed complexity if and only if $\Pi_u \in \mathsf{ZPP}$.*

Unfortunately, Theorem 2 cannot be generalized to classes of ILPs with at least one perturbed covering constraint. However, it is still true if all constraints are perturbed covering constraints.

Theorem 3. *A class Π of ILPs consisting only of perturbed covering constraints has polynomial smoothed complexity if and only if $\Pi_u \in \mathsf{ZPP}$.*

This characterization shows that strong NP-hard classes like general packing or covering ILPs do not have polynomial smoothed complexity, unless $\mathsf{ZPP} = \mathsf{NP}$. On the other hand, packing and covering problems with a fixed number of constraints like, e.g., the MAXIMUM INTEGER (k-DIMENSIONAL) KNAPSACK PROBLEM have polynomial smoothed complexity as they admit pseudopolynomial time algorithms. The same is true for ILPs with polynomially bounded $|\mathcal{D}|$ and a fixed number of constraints.

Technical comparison to previous work. In this paper we present a generalization of the smoothed analysis for binary optimization problems presented in [5] towards integer optimization problems. The rough course of the probabilistic analysis presented in the subsequent sections is similar to the analysis from [5]: We prove certain structural properties which are then exploited algorithmically in form of an adaptive rounding scheme using pseudopolynomial algorithms as a subroutine. In particular, we present a probabilistic analysis showing that it is sufficient to reveal only a logarithmic number of bits of each stochastic number in order to determine the optimal solution. We want to remark, however, that the generalization of this result from the binary to the integer case is not straightforward but technically difficult in several aspects. The major challenge we have to tackle is that the main part of the previous probabilistic analysis heavily relies on the fact that variables are binary. For example, the previous analysis uses the existence of 0 entries in any solution (except 1^n) in order to place assumptions on subsets of solutions sharing a 0 at the same position. Observe that assumptions on the values of the solutions in such subsets do not effect the random coefficients at which all these solutions take the value 0. Obviously, this elementary trick fails already when going from binary to tertiary variables. In this paper, we use a different kind of analysis that places assumptions on subsets of solutions in such a way that only values of linear combinations of pairs of random coefficients are revealed. In the subsequent analysis, the knowledge about these linear combinations is taken into account carefully.

2. Probabilistic analysis of ILPs

In order to prepare the proof of Theorem 1, we will analyze structural properties of semi-random ILPs. Let $I = (A, b, c)$ be an ILP with n integer variables $x_1, \ldots, x_n \in \mathcal{D}$ which has been generated according to the semi-random input model described above.

Throughout this analysis let $m = |\mathcal{D}|$, $m_{\max} = \{|a| \mid a \in \mathcal{D}\}$ and let $[n]$ denote the set $\{1, \ldots, n\}$. Note that $m \leq 2m_{\max} + 1$ holds for every $\mathcal{D}$.

2.1. Winner gap

In this section we assume only the coefficients in the objective function to be randomly perturbed. The constraints can be chosen arbitrarily. In fact, it is not even necessary that they are linear. In the following we assume that an arbitrary set $\mathcal{S}$ of feasible solutions is given.

The first structural property that we are interested in is the distance between the best and the second best solution. The best solution will be called the *winner* and denoted by x^*, i.e. $x^* = \operatorname{argmax}\{c^T x | x \in \mathcal{S}\}$. The second best solution will be denoted by x^{**}, i.e. $x^{**} = \operatorname{argmax}\{c^T x | x \in \mathcal{S} \setminus \{x^*\}\}$. We define the *winner gap* $\varDelta$ to be the difference between the objective values of x^* and x^{**}, that is,

$$\varDelta = c^T x^* - c^T x^{**}.$$

We will prove the following lower bound on the size of the winner gap. This proof follows roughly the line of arguments used in [5] to bound the size of the winner gap in binary optimization problems.

Lemma 1 (Isolating Lemma). *For all* $\varepsilon \geq 0$ *it holds* $\mathbf{Pr}\left[\varDelta \leq \varepsilon\right] \leq \varepsilon \cdot \phi n m^2$.

Proof. If a variable x_i takes the same value in all feasible solutions then this variable does not affect the winner gap and thus we can assume without loss of generality that for every $i \in [n]$ there exist two feasible solutions that differ in the i-th position. Under this assumption, we can define a winner gap Δ_i with respect to position i for each $i \in [n]$ by

$$\Delta_i = c^T x^* - c^T y, \text{ where } y = \text{argmax}\{c^T x | x \in \mathcal{S} \wedge x_i \neq x_i^*\}$$

denotes the best solution that differs from the winner x^* in the i-th position. In words, Δ_i is the difference between the objective value of the winner x^* and the value of a solution y that is best among those solutions that differ in the i-th bit from x^*.

Clearly, the best solution $x^* = (x_1^*, \ldots, x_n^*)$ and the second best solution $x^{**} = (x_1^{**}, \ldots, x_n^{**})$ differ in at least one position, that is, there exists an $i \in [n]$ such that $x_i^* \neq x_i^{**}$. If the best and the second best solution differ in position i then $\Delta = \Delta_i$. Thus, Δ is guaranteed to take a value also taken by at least one of the variables $\Delta_1, \ldots, \Delta_n$. Thus, we conclude

$$\mathbf{Pr}\left[\Delta \leq \varepsilon\right] \leq \sum_{i=1}^{n} \mathbf{Pr}\left[\Delta_i \leq \varepsilon\right]. \tag{2}$$

Fix an arbitrary index $i \in [n]$. We will prove an upper bound on the probability that Δ_i does not exceed ε. For that purpose, we partition the set of feasible solutions $\mathcal{S}$ into disjoint subsets $\mathcal{S}_j$ for $j \in \mathcal{D}$. We define the subset $\mathcal{S}_j$ to be the set of all feasible solutions which take value j at position i, i.e. $\mathcal{S}_j = \{x \in \mathcal{S} | x_i = j\}$. Now suppose all random variables c_k with $k \neq i$ are fixed arbitrarily. Obviously, under this assumption, the winner among the solutions in $\mathcal{S}_0$ and its objective value are fixed as the objective values of the solutions in $\mathcal{S}_0$ do not depend on c_i. Although the objective values of the solutions in $\mathcal{S}_j$, for $j \neq 0$, are not fixed, the winner of $\mathcal{S}_j$ is determined as well because the unknown outcome of the random variable c_i does not affect the order among the solutions in $\mathcal{S}_j$. For $j \in \mathcal{D}$, let $x^{(j)}$ denote a winner among the solutions in $\mathcal{S}_j$.

Since the winner x^* and the best solution differing in the i-th position from x^* cannot be contained in the same set $\mathcal{S}_j$ it always exist $j_1, j_2 \in \mathcal{D}$, $j_1 \neq j_2$, with $\Delta_i = c^T x^{(j_1)} - c^T x^{(j_2)}$. Hence, Δ_i is always equal to one of the random variables $c^T x^{(j_1)} - c^T x^{(j_2)}$ with $j_1, j_2 \in \mathcal{D}$, $j_1 \neq j_2$ and therefore

$$\mathbf{Pr}\left[\Delta_i \leq \varepsilon\right] \leq \sum_{\substack{j_1, j_2 \in \mathcal{D} \\ j_1 \neq j_2}} \mathbf{Pr}\left[c^T x^{(j_1)} - c^T x^{(j_2)} \in [0, \varepsilon]\right]. \tag{3}$$

Observe that, under the assumption that the c_k with $k \neq i$ are fixed, the random variable $c^T x^{(j_1)} - c^T x^{(j_2)}$ can be rewritten as $\kappa + (j_1 - j_2)c_i$ for some constant κ depending on the c_k with $k \neq i$. Since we assumed the density of c_i to be bounded by ϕ the density of $\kappa + (j_1 - j_2)c_i$ is bounded by $\phi/(|j_1 - j_2|) \leq \phi$. Hence, for any $j_1, j_2 \in \mathcal{D}$ with $j_1 \neq j_2$, we have

$$\mathbf{Pr}\left[c^T x^{(j_1)} - c^T x^{(j_2)} \in [0, \varepsilon]\right] \leq \varepsilon \cdot \phi. \tag{4}$$

Since there are less than m^2 different possible choices for j_1 and j_2 combining (3) and (4) yields

$$\mathbf{Pr}\left[\Delta_i \leq \varepsilon\right] \leq \varepsilon \cdot \phi m^2$$

and therefore with (2) we obtain

$$\mathbf{Pr}\left[\Delta \leq \varepsilon\right] \leq \varepsilon \cdot \phi n m^2.$$

□

2.2. Loser and feasibility gap for a single constraint.

Now we assume the coefficients in the objective function to be fixed arbitrarily. Solutions with the same objective values are ranked in an arbitrary but fixed fashion. In this section we deal only with the case $k = 1$, that is the set of feasible solutions is determined by one linear constraint. For this case, we will define two structural properties called *loser gap* and *feasibility gap* and prove lower bounds for these gaps holding with a certain probability. A generalization to the general case with $k > 1$ many constraints will be given in the next section.

We assume that the constraint determining the set of feasible solutions is of the form $w^T x = w_1 x_1 + \ldots + w_n x_n \leq t$ where the coefficients $w_1, \ldots, w_n$ correspond to independent random variables following possibly different probability distributions with bounded densities $f_1, \ldots, f_n$, respectively. For technical reasons, we have to allow further restrictions on the set of feasible solutions. To be more precise, we assume that an arbitrary subset $\mathcal{S} \subseteq \mathcal{D}^n$ is given and that the set of feasible solutions is obtained as intersection of $\mathcal{S}$ with the half-space $\mathcal{B}$ described by the constraint $w^T x \leq t$. For $i \in [n]$, let $\phi_i = \sup_{s \in \mathbb{R}} f_i(s)$ and $\phi = \max_{i \in [n]} \phi_i$. The *winner*, denoted by x^*, is the solution with highest rank in $\mathcal{S} \cap \mathcal{B}$. The *feasibility gap* is defined by

$$\varGamma = \begin{cases} t - w^T x^* & \text{if } \mathcal{S} \cap \mathcal{B} \neq \emptyset \\ \perp & \text{otherwise} \end{cases} .$$

In words, $\varGamma$ corresponds to the slack of the winner with respect to the threshold t. A solution from $\mathcal{S}$ is called a *loser* if it has a higher rank than x^*, that is, the losers are those solutions from $\mathcal{S}$ that are better than the winner (w.r.t. the ranking) but that are cut off by the constraint $w^T x \leq t$. The set of losers is denoted by $\mathcal{L}$. If there is no winner, as there is no feasible solution, then we define $\mathcal{L} = \mathcal{S}$. The *loser gap* is defined by

$$\varLambda = \begin{cases} \min\{w^T x - t | x \in \mathcal{L}\} & \text{if } \mathcal{L} \neq \emptyset \\ \perp & \text{otherwise} \end{cases} .$$

Our goal is to show that both the loser and the feasibility gap of a semi-random ILP are lower bounded by a polynomial in $(nm_{\max}\phi)^{-1}$ with probability close to 1.

Observe that the solution 0^n is different from all other solutions in $\mathcal{S}$ as its feasibility does not depend on the outcome of the random coefficients $w_1, \ldots, w_n$. Suppose $0^n \in \mathcal{S}$ and 0^n has the highest rank among all solutions in $\mathcal{S}$. Then one can enforce $\Gamma = 0$ by setting $t = 0$. Similarly, one can enforce $\Lambda \to 0$ for $t < 0$ and $t \to 0$. For this reason, we need to exclude the solution 0^n from our analysis. Later we will describe how the random perturbation of the threshold helps us to cope with this problem.

The key result of this section is the following lemma about the sizes of loser and feasibility gap.

Lemma 2 (Separating Lemma). *Let $\mathcal{S}$ with $0^n \notin \mathcal{S}$ be chosen arbitrarily and let $c = \max_{i\in[n]} \mathbf{E}\left[|w_i|\right]$. Then, for all ε with $\varepsilon \leq (32n^5m^7m_{\max}\phi^2)^{-1}$, $\mathbf{Pr}\left[\Gamma \leq \varepsilon\right] \leq 2(\varepsilon \cdot 32cn^5m^7m_{\max}\phi^2)^{1/3}$ and $\mathbf{Pr}\left[\Lambda \leq \varepsilon\right] \leq 2(\varepsilon \cdot 32cn^5m^7m_{\max}\phi^2)^{1/3}$.*

The proof of this lemma is subdivided into a few steps. At first, we will assume that the densities $f_1, \ldots, f_n$ have a bounded support, i.e. we assume the existence of a constant $s \in \mathbb{R}_{\geq 0}$ such that $f_i(x) = 0$ holds for every $i \in [n]$ and for every $x \notin [-s, s]$. In addition to that, we assume that the set $\mathcal{S}$ does only contain elements which are pairwise linearly independent, i.e. we assume that there do not exist two solutions $x, y \in \mathcal{S}$ such that $x = \alpha y$ or $y = \alpha x$ holds for some $\alpha \in \mathbb{R}$. In this case we can show an upper bound on the probability that the loser gap does not exceed ε. Then, we will use symmetry properties between the two gaps in order to show that bounds for the loser gap also hold for the feasibility gap and vice versa. Thus, the bound proven for the loser gap holds for the feasibility gap as well. The assumption that the set $\mathcal{S}$ does not contain linearly dependent solutions can be dropped at the cost of an extra factor m for the feasibility gap. Due to the symmetry this bound also applies to the loser gap. In the last step we will drop the assumption that the support of the densities is bounded.

Lemma 3. *Let $\mathcal{S}$ with $0^n \notin \mathcal{S}$ be chosen arbitrarily such that $\mathcal{S}$ does not contain two linearly dependent solutions. Assume $f_i(x) = 0$, for $i \in [n]$ and $x \notin [-s, s]$. Then, for all $\varepsilon \geq 0$ and for all $p \geq 1$,*

$$\mathbf{Pr}\left[\varLambda \leq \varepsilon\right] \leq \frac{1}{2p} + \varepsilon \cdot 4n^4 m^6 m_{\max} \phi^2 sp.$$

The role of the parameter p needs some explanation. We will show that the density of the loser gap is upper bounded by $4n^4 m^6 m_{\max} \phi^2 sp$ if some failure event $\mathcal{E}(p)$ does not occur. It holds $\mathbf{Pr}\left[\mathcal{E}(p)\right] \leq 1/(2p)$. Thus, the first addend corresponds to the case of failure $\mathcal{E}(p)$ and the second one corresponds to the case $\neg\mathcal{E}(p)$. Note that an upper bound α on the density of $\varLambda$ implies an upper bound on the probability that $\varLambda$ takes a value less than or equal to ε of $\varepsilon\alpha$ since $\varLambda$ takes only non-negative values.

Now we will present our approach to bound the density $f_\varLambda$ of the random variable $\varLambda$. We will see that this approach fails under certain circumstances and define the failure event $\mathcal{E} := \mathcal{E}(p)$ accordingly. For each combination of $i, j \in [n]$ with $i < j$ and of $\overline{m} = (m_1, m_2, m_3, m_4) \in \mathcal{D}^4$ with linearly independent vectors (m_1, m_2) and (m_3, m_4), we define a random variable $\varLambda_{i,j}^{\overline{m}}$ in such a way that there are always indices i, j and a vector $\overline{m}$ such that $\varLambda = \varLambda_{i,j}^{\overline{m}}$ holds. Hence, it follows

$$\mathbf{Pr}\left[\varLambda \leq \varepsilon\right] = \mathbf{Pr}\left[\varLambda \in [0, \varepsilon]\right] \leq \sum_{i,j,\overline{m}} \mathbf{Pr}\left[\varLambda_{i,j}^{\overline{m}} \in [0, \varepsilon]\right].$$

For this reason, a bound on the densities of the random variables $\varLambda_{i,j}^{\overline{m}}$ implies a bound on the probability that $\varLambda$ does not exceed ε.

Let x^* denote the winner, let $x^{\min}$ denote the *minimal loser*, i.e. $x^{\min} = \operatorname{argmin}\{w^T x | x \in \mathcal{L}\}$, and fix some $i, j \in [n]$ with $i < j$ and a vector $\overline{m} \in \mathcal{D}^4$ with linearly independent subvectors (m_1, m_2) and (m_3, m_4). First of all, we will formally define the random variable $\Lambda_{i,j}^{\overline{m}}$. Therefore, let $x_{i,j}^{*,m_3,m_4}$ denote the winner of those solutions x with $x_i = m_3$ and $x_j = m_4$, i.e. $x_{i,j}^{*,m_3,m_4}$ denotes the highest ranked solution in $\{x \in \mathcal{S} | x_i = m_3, x_j = m_4\} \cap \mathcal{B}$. Based on this definition, we define a set of losers

$$\mathcal{L}_{i,j}^{\overline{m}} = \{x \in \mathcal{S} \mid x_i = m_1, x_j = m_2,\ x \text{ is ranked higher than } x_{i,j}^{*,m_3,m_4}\}.$$

The minimal loser $x_{i,j}^{\min,\overline{m}}$ is defined to be the solution from $\mathcal{L}_{i,j}^{\overline{m}}$ with the smallest weight, i.e. $x_{i,j}^{\min,\overline{m}} = \operatorname{argmin}\{w^T x \mid x \in \mathcal{L}_{i,j}^{\overline{m}}\}$. Now the random variable $\Lambda_{i,j}^{\overline{m}}$ is defined to be the slack of the minimal loser $x_{i,j}^{\min,\overline{m}}$ w.r.t. the threshold t, i.e. $\Lambda_{i,j}^{\overline{m}} = w^T x_{i,j}^{\min,\overline{m}} - t$. If $\mathcal{L}_{i,j}^{\overline{m}} = \emptyset$ then $x_{i,j}^{\min,\overline{m}}$ and $\Lambda_{i,j}^{\overline{m}}$ are undefined.

One can easily argue that the requirement that Λ always takes a value equal to one of the values of the $\Lambda_{i,j}^{\overline{m}}$ is fulfilled: The winner x^* and the minimal loser $x^{\min}$ are linearly independent since they are both elements from $\mathcal{S}$. Thus, there are always two indices $i, j \in [n]$ with $i < j$ such that the vectors (x_i^*, x_j^*) and $(x_i^{\min}, x_j^{\min})$ are linearly independent. Setting $(m_3, m_4) = (x_i^*, x_j^*)$ and $(m_1, m_2) = (x_i^{\min}, x_j^{\min})$ yields $\Lambda = \Lambda_{i,j}^{\overline{m}}$.

In order to obtain an upper bound on the density of the random variable $\Lambda_{i,j}^{\overline{m}}$, we reduce the degree of randomness. That is, we assume the coefficients w_k with $k \neq i$ and $k \neq j$ and the sum $m_3 w_i + m_4 w_j$ to be fixed arbitrarily. An upper bound for the density of $\Lambda_{i,j}^{\overline{m}}$ holding for all deterministic choices of these random variables obviously holds for all random choices as well. The winner $x_{i,j}^{*,m_3,m_4}$ can be determined without knowing the outcome of w_i and w_j as the weights of all solutions in $\{x \in \mathcal{S} | x_i = m_3, x_j = m_4\}$ are known. Thus, also $\mathcal{L}_{i,j}^{\overline{m}}$ is known. Since the random variables w_i and w_j affect the weight of all solutions in $\mathcal{L}_{i,j}^{\overline{m}}$ in the same fashion, also the minimal loser $x_{i,j}^{\min,\overline{m}}$ does not depend on the outcome of w_i and w_j. Hence, if the outcome of w_k with $k \neq i$ and $k \neq j$ and the sum $m_3 w_i + m_4 w_j$ are known, the loser gap $\Lambda_{i,j}^{\overline{m}}$ can be rewritten as

$$\Lambda_{i,j}^{\overline{m}} = w^T x_{i,j}^{\min,\overline{m}} - t = \kappa + m_1 w_i + m_2 w_j ,$$

where κ denotes a constant depending on the fixed values of w_k with $k \neq i$ and $k \neq j$ and $m_3 w_i + m_4 w_j$. Thus, under our assumption, $\Lambda_{i,j}^{\overline{m}}$ and $m_1 w_i + m_2 w_j$ are random variables which differ only by a constant addend. In particular, upper bounds on the density of the random variable $m_1 w_i + m_2 w_j$ hold for the density of $\Lambda_{i,j}^{\overline{m}}$ as well. Recall that we still assume the sum $m_3 w_i + m_4 w_j$ to be fixed to an arbitrary value $z \in \mathbb{R}$. Therefore, we will determine the conditional density $g_{i,j}^{\overline{m},z}$ of $m_1 w_i + m_2 w_j$ under the condition $m_3 w_i + m_4 w_j = z$.

Lemma 4. *Let $(m_1, m_2) \in \mathcal{D}^2$ and $(m_3, m_4) \in \mathcal{D}^2$ be linearly independent and let $f_{m_3 w_i + m_4 w_j} : \mathbb{R} \to \mathbb{R}$ denote the density of the random variable $m_3 w_i + m_4 w_j$. Furthermore, let $g_{i,j}^{\overline{m},z} : \mathbb{R} \to \mathbb{R}$ denote the conditional density of the random variable $m_1 w_i + m_2 w_j$ under the condition $m_3 w_i + m_4 w_j = z$. Then*

$$g_{i,j}^{\overline{m},z}(x) \leq \frac{\phi^2}{f_{m_3 w_i + m_4 w_j}(z)} . \tag{5}$$

Proof. Let $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}_{\geq 0}$ denote the joint density of the random variables $A = m_1 w_i + m_2 w_j$ and $B = m_3 w_i + m_4 w_j$. Since the vectors (m_1, m_2) and (m_3, m_4) are assumed to be linearly independent, the transformation $\Phi : \mathbb{R}^2 \to \mathbb{R}^2$ with $\Phi(x, y) =$

$(m_1x + m_2y, m_3x + m_4y)$ is bijective and can be inverted as follows

$$\Phi^{-1}(a,b) = \left(\frac{m_4a - m_2b}{m_1m_4 - m_2m_3}, \frac{m_1b - m_3a}{m_1m_4 - m_2m_3}\right).$$

In order to determine the conditional density $g_{i,j}^{\overline{m},z}$, we have to determine the Jacobian matrix M of the transformation Φ^{-1} containing the partial derivatives of Φ^{-1} as matrix entries. With $d = m_1m_4 - m_2m_3$ we have

$$M = \begin{pmatrix} \frac{m_4}{d} & -\frac{m_2}{d} \\ -\frac{m_3}{d} & \frac{m_1}{d} \end{pmatrix}.$$

The determinant of the Jacobian matrix is $1/d$. Due to the independence of the random variables w_i and w_j, the joint density f of A and B can be written as

$$\begin{aligned} f(a,b) &= |\det M| \cdot f_i(\Phi_1^{-1}(a,b)) \cdot f_j(\Phi_2^{-1}(a,b)) \\ &= \frac{1}{|d|} \cdot f_i\left(\frac{m_4a - m_2b}{d}\right) \cdot f_j\left(\frac{m_1b - m_3a}{d}\right) \\ &\le \frac{\phi^2}{|d|} \le \phi^2. \end{aligned}$$

The conditional density $g_{i,j}^{\overline{m},z}$ can be expressed as follows

$$g_{i,j}^{\overline{m},z}(x) = \frac{f(x,z)}{\int_{\mathbb{R}} f(x,z)\,dx} = \frac{f(x,z)}{f_{m_3w_i+m_4w_j}(z)}.$$

Thus, for all $x \in \mathbb{R}$, it holds

$$g_{i,j}^{\overline{m},z}(x) \le \frac{\phi^2}{f_{m_3w_i+m_4w_j}(z)}.$$

□

Hence, $g_{i,j}^{\overline{m},z}$ cannot be upper bounded since, in general, the denominator in (5) can become arbitrarily small. Therefore, we have to restrict the possible choices for z to the set $\mathbb{R}\backslash M_{i,j}^{m_3,m_4}$ with

$$M_{i,j}^{m_3,m_4} = \left\{ z \in \mathbb{R} \,\middle|\, 0 \le f_{m_3 w_i + m_4 w_j}(z) \le \frac{1}{4n^2 m^2 m_{\max} sp} \right\}.$$

We will denote the event that $m_3 w_i + m_4 w_j$ takes a value from $M_{i,j}^{m_3,m_4}$ by $\mathcal{E}_{i,j}^{m_3,m_4}$. In case of $\neg\mathcal{E}_{i,j}^{m_3,m_4}$ the conditional density $g_{i,j}^{\overline{m},z}$ is bounded from above by $4n^2 m^2 m_{\max} sp\phi^2$. Hence, $\mathbf{Pr}\left[\Lambda_{i,j}^{\overline{m}} \in [0,\varepsilon] | \neg\mathcal{E}_{i,j}^{m_3,m_4}\right] \le \varepsilon \cdot 4n^2 m^2 m_{\max} \phi^2 sp$. In the following lemma we will use the bounded support of the densities f_i and f_j to show that it is unlikely that the event $\mathcal{E}_{i,j}^{m_3,m_4}$ occurs. Let $\mathcal{E}$ denote the event that, for at least one combination of $i, j \in [n]$ and $m_3, m_4 \in \mathcal{D}$, the event $\mathcal{E}_{i,j}^{m_3,m_4}$ occurs, that is, $\mathcal{E}$ denotes the union of all these events.

Lemma 5. *It holds*

$$\mathbf{Pr}\left[\Lambda \le \varepsilon | \neg\mathcal{E}\right] \le \varepsilon \cdot 4n^4 m^6 m_{\max} \phi^2 sp$$

and

$$\mathbf{Pr}\left[\mathcal{E}\right] \le \frac{1}{2p}.$$

Proof. We have defined $\mathcal{E}_{i,j}^{m_3,m_4}$ to be the event that the random variable $m_3 w_i + m_4 w_j$ takes a value from the set

$$M_{i,j}^{m_3,m_4} = \left\{ z \in \mathbb{R} \,\middle|\, 0 \le f_{m_3 w_i + m_4 w_j}(z) \le \frac{1}{4n^2 m^2 m_{\max} sp} \right\}.$$

The probability of this event can be written as follows

$$\mathbf{Pr}\left[\mathcal{E}_{i,j}^{m_3,m_4}\right] = \int_{M_{i,j}^{m_3,m_4}} f_{m_3 w_i + m_4 w_j}(z)\, dz.$$

We define

$$M_{i,j}^{*,m_3,m_4} = \left\{ z \in M_{i,j}^{m_3,m_4} \,\middle|\, f_{m_3 w_i + m_4 w_j}(z) > 0 \right\}$$

and obtain the following estimate

$$\begin{aligned}\mathbf{Pr}\left[E_{i,j}^{m_3,m_4}\right] &= \int_{M_{i,j}^{m_3,m_4}} f_{m_3w_i+m_4w_j}(z)\,dz \\ &\leq \frac{1}{4n^2m^2m_{\max}sp}\int_{M_{i,j}^{*,m_3,m_4}} 1\,dz. \end{aligned} \tag{6}$$

We will prove an upper bound of $4m_{\max}s$ on the integral occurring in (6) in order to obtain the desired bound on the probability of the event $\mathcal{E}_{i,j}^{m_3,m_4}$.

1st case: $m_3 = 0$ and $m_4 \neq 0$. In this case, it holds $f_{m_3w_i+m_4w_j} = f_{m_4w_j}$, where $f_{m_4w_j}$ denotes the density of the random variable m_4w_j. We obtain

$$\begin{aligned}M_{i,j}^{*,m_3,m_4} &\subseteq \left\{z \in \mathbb{R} \,\middle|\, f_{m_4w_j}(z) > 0\right\} \\ &= \left\{z \in \mathbb{R} \,\middle|\, \frac{1}{|m_4|}\cdot f_j\left(\frac{z}{m_4}\right) > 0\right\} \\ &\subseteq \left\{z \in \mathbb{R} \,\middle|\, -s \leq \frac{z}{m_4} \leq s\right\} \\ &= [-s|m_4|, s|m_4|].\end{aligned}$$

Altogether, we obtain

$$\int_{M_{i,j}^{*,m_3,m_4}} 1\,dz \leq 2s\,|m_4| \leq 2m_{\max}s,$$

2nd case: $m_3 \neq 0$ and $m_4 = 0$. Analogous to the first case.

Preparation of the following cases. In the cases which we have not yet considered it holds $m_3 \neq 0$ and $m_4 \neq 0$. Therefore, the density $f_{m_3w_i+m_4w_j}(z)$ can be rewritten as follows:

$$\begin{aligned} f_{m_3 w_i + m_4 w_j}(z) &= \int_{-\infty}^{\infty} f_{m_3 w_i}(x) f_{m_4 w_j}(z-x)\, dx \\ &= \frac{1}{|m_3| \cdot |m_4|} \cdot \int_{-\infty}^{\infty} f_i\left(\frac{x}{m_3}\right) f_j\left(\frac{z-x}{m_4}\right) dx \\ &= \frac{1}{|m_4|} \cdot \int_{-\infty}^{\infty} f_i\left(x\right) \cdot f_j\left(\frac{z-m_3 x}{m_4}\right) dx. \end{aligned}$$

Thus, for $m_3 \neq 0$ and $m_4 \neq 0$, it holds

$$\begin{aligned} M_{i,j}^{*,m_3,m_4} &\subseteq \left\{ z \in \mathbb{R} \,\middle|\, \frac{1}{|m_4|} \cdot \int_{-\infty}^{\infty} f_i\left(x\right) \cdot f_j\left(\frac{z-m_3 x}{m_4}\right) dx > 0 \right\} \\ &\subseteq \left\{ z \in \mathbb{R} \,\middle|\, \exists x \in \mathbb{R} : f_i\left(x\right) \cdot f_j\left(\frac{z-m_3 x}{m_4}\right) > 0 \right\} \\ &= \left\{ z \in \mathbb{R} \,\middle|\, \exists x \in \mathbb{R} : f_i\left(x\right) > 0 \wedge f_j\left(\frac{z-m_3 x}{m_4}\right) > 0 \right\} \\ &\subseteq \left\{ z \in \mathbb{R} \,\middle|\, \exists x \in \mathbb{R} : \left(-s \leq x \leq s\right) \wedge \left(-s \leq \frac{z-m_3 x}{m_4} \leq s\right) \right\}. \end{aligned} \tag{7}$$

3rd case: $m_3 \neq 0, m_4 \neq 0$ and $m_3 \cdot m_4 > 0$. We start by rewriting the second inequality in (7). We have

$$-s \leq \frac{z-m_3 x}{m_4} \leq s \iff -\frac{m_4}{m_3}s + \frac{z}{m_3} \leq x \leq \frac{m_4}{m_3}s + \frac{z}{m_3}.$$

Hence, the inequalities in (7) imply the following lower bounds for x

$$x_{l,1} = -s \text{ and } x_{l,2} = -\frac{m_4}{m_3}s + \frac{z}{m_3}$$

and the following upper bounds for x

$$x_{u,1} = s \text{ and } x_{u,2} = \frac{m_4}{m_3}s + \frac{z}{m_3}.$$

For any given z, the domain of the variable x is restricted by these bounds to the interval $I = [\max\{x_{l,1}, x_{l,2}\}, \min\{x_{u,1}, x_{u,2}\}]$. If this interval is empty then z does not belong to the set $M_{i,j}^{*,m_3,m_4}$. In order to determine the values of z which yield $I = \emptyset$ we solve the equations $x_{l,1} = x_{u,2}$ and $x_{l,2} = x_{u,1}$ w.r.t. z. We obtain

$$x_{l,1} = x_{u,2} \iff z = -(m_3 + m_4)s$$

and

$$x_{l,2} = x_{u,1} \iff z = (m_3 + m_4)s.$$

Subcase 3a: $m_3 > 0$ **and** $m_4 > 0$**.** In this subcase it holds

$$\begin{aligned} z &< -(m_3 + m_4)s \Rightarrow x_{l,1} > x_{u,2} \text{ and} \\ z &> (m_3 + m_4)s \;\;\Rightarrow x_{l,2} > x_{u,1}. \end{aligned}$$

Thus, setting $z < -(m_3 + m_4)s$ or $z > (m_3 + m_4)s$ yields $I = \emptyset$. Altogether, we obtain

$$M_{i,j}^{*,m_3,m_4} \subseteq [-(m_3 + m_4)s, (m_3 + m_4)s]$$

and, therefore,

$$\int_{M_{i,j}^{*,m_3,m_4}} 1 \, db \leq 2(m_3 + m_4)s \leq 4m_{\max}s.$$

Subcase 3b: $m_3 < 0$ **and** $m_4 < 0$**.** Analogous to case 3a.

4th case: $m_3 \neq 0, m_4 \neq 0$ **and** $m_3 \cdot m_4 < 0$**.** Analogous to the third case.

Substituting the integral in (6) by the estimate $4m_{\max}s$ yields

$$\mathbf{Pr}\left[\mathcal{E}_{i,j}^{m_3,m_4}\right] \leq \frac{1}{n^2m^2p}.$$

Hence, it holds

$$\mathbf{Pr}\left[\mathcal{E}\right] = \mathbf{Pr}\left[\bigcup_{i,j,m_3,m_4} \mathcal{E}_{i,j}^{m_3,m_4}\right] \leq \binom{n}{2} m^2 \cdot \frac{1}{n^2 m^2 p} \leq \frac{1}{2p}.$$

We obtain

$$\begin{aligned}
\mathbf{Pr}\left[\Lambda_{i,j}^{\overline{m}} \in [0,\varepsilon] | \neg\mathcal{E}\right] &= \frac{\mathbf{Pr}\left[\Lambda_{i,j}^{\overline{m}} \in [0,\varepsilon] \wedge \neg\mathcal{E}\right]}{\mathbf{Pr}\left[\neg\mathcal{E}\right]} \\
&\leq \frac{1}{1-1/(2p)} \cdot \mathbf{Pr}\left[\Lambda_{i,j}^{\overline{m}} \in [0,\varepsilon] \wedge \neg\mathcal{E}_{i,j}^{m_3,m_4}\right] \\
&\leq 2 \cdot \mathbf{Pr}\left[\Lambda_{i,j}^{\overline{m}} \in [0,\varepsilon] | \neg\mathcal{E}_{i,j}^{m_3,m_4}\right] \\
&\leq \varepsilon \cdot 8n^2 m^2 m_{\max} \phi^2 sp.
\end{aligned}$$

Hence, we obtain the desired result:

$$\begin{aligned}
&\mathbf{Pr}\left[\Lambda \leq \varepsilon | \neg\mathcal{E}\right] \\
&\qquad \leq \sum_{i,j,\overline{m}} \mathbf{Pr}\left[\Lambda_{i,j}^{\overline{m}} \in [0,\varepsilon] | \neg\mathcal{E}\right] \\
&\qquad \leq \varepsilon \cdot \binom{n}{2} m^4 \cdot 8n^2 m^2 m_{\max} \phi^2 sp \\
&\qquad \leq \varepsilon \cdot 4n^4 m^6 m_{\max} \phi^2 sp.
\end{aligned}$$

□

Proof (Lemma 3). Now Lemma 3 follows as an easy corollary from Lemma 5:

$$\mathbf{Pr}\left[\Lambda \leq \varepsilon\right] \leq \mathbf{Pr}\left[\mathcal{E}\right] + \mathbf{Pr}\left[\Lambda \leq \varepsilon | \neg\mathcal{E}\right] \leq \frac{1}{2p} + \varepsilon \cdot 4n^4 m^6 m_{\max} \phi^2 sp.$$

□

Now we show that Lemma 3 holds for the feasibility gap as well. First of all, we have to generalize the definitions of loser and feasibility gap a little bit. Let $\Lambda(t)$ denote the loser gap w.r.t. the constraint $w^T x \leq t$ and let $\Gamma(t)$ denote the feasibility gap w.r.t. this constraint.

Lemma 6. *Let $t \in \mathbb{R}$ and $\varepsilon \in \mathbb{R}_{\geq 0}$ arbitrary, then* $\mathbf{Pr}\left[\Lambda(t) \leq \varepsilon | \neg\mathcal{E}\right] = \mathbf{Pr}\left[\Gamma(t+\varepsilon) \leq \varepsilon | \neg\mathcal{E}\right]$.

This lemma can be proven by arguments very similar to those used in the proof of Lemma 9 in [5]. For the sake of completeness, we provide the proof here.

Proof (Lemma 6).

We take an alternative view on the given optimization problem. We interpret the problem as a bicriteria problem. The feasible region is defined by the set $\mathcal{S}$. On the one hand, we seek for a solution from $\mathcal{S}$ whose rank is as high as possible. On the other hand, we seek for a solution with small weight, where the *weight* of a solution $x \in \mathcal{S}$ is defined by the linear function $w^T x$. A solution $x \in \mathcal{S}$ is called *Pareto-optimal* if there is no higher ranked solution $y \in \mathcal{S}$ with weight at most $w^T x$. Let $\mathcal{P}$ denote the set of Pareto-optimal solutions.

Next we show that winners and minimal losers of the original optimization problem correspond to Pareto-optimal solutions of the bicriteria problem. First, let us observe that the winner x^* with respect to any given weight threshold t is a Pareto-optimal solution for the bicriteria problem because there is no other solution with a higher rank and a smaller weight than $t \geq w^T x^*$. Moreover, for every Pareto-optimal solution x there is also a threshold t such that x is the winner, i.e. $t = w^T x$.

The same kind of characterization holds for minimal losers as well. Recall, for a given threshold t, the minimal loser is defined to be $x^{\min} = \operatorname{argmin}\{w^T x | x \in \mathcal{L}\}$. We claim that there is no other solution y that simultaneously achieves a higher rank and not larger weight than $x^{\min}$. This can be seen as follows. Suppose y is a solution with higher rank than $x^{\min}$. If $w^T y \leq t$ then $y \in \mathcal{B}$ and, hence, $x^{\min}$ would not be a loser. However, if $w^T y \in (t, w^T x^{\min}]$ then y and $x^{\min}$ would both be losers, but y instead of $x^{\min}$ would be minimal. Here we implicitly assume that there are no two solutions with the same weight. This assumption is justified as the probability that there are two solutions with the same weight is 0. Furthermore, for every Pareto-optimal solution x there is also a threshold t such that x is a minimal loser. This threshold can be obtained by setting $t \to w^T x$, $t < w^T x$.

Now let us describe loser and feasibility gap in terms of Pareto-optimal solutions. Let $\mathcal{P} \subseteq \mathcal{S}$ denote the set of Pareto-optimal solutions with respect to the fixed ranking and the random weight function $w^T x$. Then feasibility and loser gap are characterized by

$$\begin{aligned} \Gamma(t) &= \min\{t - w^T x | x \in \mathcal{P}, w^T x \leq t\}, \\ \Lambda(t) &= \min\{w^T x - t | x \in \mathcal{P}, w^T x > t\}. \end{aligned}$$

For a better intuition, we can imagine that all Pareto-optimal solutions are mapped onto a horizontal line such that a Pareto-optimal solution x is mapped to the point $w^T x$. Then $\Gamma(t)$ is the distance from the point t on this line to the closest Pareto point left to t (i.e. less than or equal to t) , and $\Lambda(t)$ is the distance from t to the closest Pareto point strictly right of t (i.e. larger than t). That is,

$$\begin{aligned} \mathbf{Pr}\left[\Lambda(t) \leq \varepsilon | \neg\mathcal{E}\right] &= \mathbf{Pr}\left[\exists x \in \mathcal{P} : w^T x \in (t, t+\varepsilon] | \neg\mathcal{E}\right] \\ &= \mathbf{Pr}\left[\Gamma(t+\varepsilon) \leq \varepsilon | \neg\mathcal{E}\right]. \end{aligned}$$

□

Corollary 1. *Let $\mathcal{S}$ with $0^n \notin \mathcal{S}$ be chosen arbitrarily such that $\mathcal{S}$ does not contain two linearly dependent solutions. Assume $f_i(x) = 0$, for $i \in [n]$ and $x \notin [-s, s]$. Then, for* *all* $\varepsilon \geq 0$ *and for all* $p \geq 1$,

$$\mathbf{Pr}\left[\Gamma \leq \varepsilon\right] \leq \frac{1}{2p} + \varepsilon \cdot 4n^4 m^6 m_{\max} \phi^2 sp.$$

$\square$

Proof. Observe that Lemma 3 and Lemma 6 hold for arbitrary choices of t. In particular, for given $t \in \mathbb{R}$ and given $\varepsilon > 0$, Lemma 3 holds for $t' = t - \varepsilon$. Hence,

$$\begin{aligned}\mathbf{Pr}\left[\Gamma \leq \varepsilon\right] &= \mathbf{Pr}\left[\Gamma(t) \leq \varepsilon\right] \\ &= \mathbf{Pr}\left[\Lambda(t-\varepsilon) \leq \varepsilon\right] \\ &\leq \frac{1}{2p} + \varepsilon \cdot 4n^4 m^6 m_{\max} \phi^2 sp.\end{aligned}$$

$\square$

Next, we will drop the assumption that the set of feasible solutions $\mathcal{S}$ does not contain linearly dependent solutions and obtain the following result.

Lemma 7. *Let* $\mathcal{S}$ *with* $0^n \notin \mathcal{S}$ *be chosen arbitrarily. Assume* $f_i(x) = 0$, *for* $i \in [n]$ *and* $x \notin [-s, s]$. *Then, for all* $\varepsilon \geq 0$ *and for all* $p \geq 1$, $\mathbf{Pr}\left[\Gamma < \varepsilon | \neg\mathcal{E}\right] \leq \varepsilon \cdot 4n^4 m^7 m_{\max} \phi^2 sp$ *and* $\mathbf{Pr}\left[\Lambda < \varepsilon | \neg\mathcal{E}\right] \leq \varepsilon \cdot 4n^4 m^7 m_{\max} \phi^2 sp$.

Proof. The main idea of the proof is to partition the set $\mathcal{S}$ into m classes $\mathcal{S}^{(1)}, \ldots, \mathcal{S}^{(m)}$ such that none of these classes contains two linearly dependent solutions. Let $\mathcal{D} = \{d_1, \ldots, d_m\}$. If $0 \notin \mathcal{D}$, such a partition can simply be created by setting $\mathcal{S}^{(k)} = \{x \in \mathcal{S} | x_1 = d_k\}$, for $k \in [m]$. Otherwise, we assume w.l.o.g. $d_m = 0$ and we set, for $k \in [m-1]$,

$$\mathcal{S}^{(k)} = \{x \in \mathcal{S} \mid \exists i \in [n] : x_1 = \ldots = x_{i-1} = 0 \text{ and } x_i = d_k\}.$$

For each of these classes a feasibility gap $\Gamma^{(k)}$ is defined. First we define the winner $x^{*,(k)}$ w.r.t. $\mathcal{S}^{(k)}$ to be that element from $\mathcal{S}^{(k)} \cap \mathcal{B}$ which is ranked highest. The feasibility gap $\Gamma^{(k)}$ is simply defined as $t - w^T x^{*,(k)}$, if $\mathcal{S}^{(k)} \neq \emptyset$, and $\perp$ otherwise. Since the winner x^* of the original problem is contained in one of the classes $\mathcal{S}^{(k)}$, the feasibility gap Γ always takes the value of one of the variables $\Gamma^{(k)}$. Observe that Lemma 3 can be applied to the subproblems defined by the classes $\mathcal{S}^{(k)}$ since these classes do not contain linearly dependent solutions. Hence, we can combine Lemma 5 and Lemma 6 to obtain $\mathbf{Pr}\left[\Gamma^{(k)} \leq \varepsilon | \neg\mathcal{E}\right] \leq \varepsilon \cdot 4n^4 m^6 m_{\max} \phi^2 sp$. Thus, it holds

$$\mathbf{Pr}\left[\Gamma \leq \varepsilon | \neg\mathcal{E}\right] \leq \sum_{k=1}^{m} \mathbf{Pr}\left[\Gamma^{(k)} \leq \varepsilon | \neg\mathcal{E}\right] \leq \varepsilon \cdot 4n^4 m^7 m_{\max} \phi^2 sp.$$

The result on the loser gap follows by another application of Lemma 6 analogously to Corollary 1. □

Now we will drop the assumption that the densities $f_1, \ldots, f_n$ have bounded supports and finish the proof of Lemma 2.

Proof (Lemma 2). The main idea is to choose some constant $s \in \mathbb{R}$ such that the probability that one of the coefficients $w_1, \ldots, w_n$ takes a value outside of the interval $[-s, s]$ is upper bounded by $1/(2p)$. We set $s = 2npc$. For $i \in [n]$, let $\mathcal{G}_i$ denote the event that $w_i \notin [-s, s]$ and let $\mathcal{G}$ denote the union of these events. An application of Markov's inequality shows $\mathbf{Pr}[\mathcal{G}] \leq 1/(2p)$. For the conditional density functions it holds

$$f_{i|\neg\mathcal{G}}(x) = \begin{cases} 0 & \text{if } x \notin [-s, s] \\ \frac{f_i(x)}{\mathbf{Pr}[w_i \in [-s,s]]} & \text{otherwise} \end{cases} \leq \begin{cases} 0 & \text{if } x \notin [-s, s] \\ 2f_i(x) & \text{otherwise} \end{cases}.$$

Thus, the densities of the random variables $w_1, \dots, w_n$ have a bounded support under the condition $\neg\mathcal{G}$. We define $\mathcal{F} = \mathcal{E} \cup \mathcal{G}$ to be the failure event. Then we can bound the probability that the loser gap or the feasibility gap does not exceed ε under the condition $\neg\mathcal{F}$. We have seen, that the condition $\neg\mathcal{G}$ leads to a conditional density which is by a factor of at most 2 larger than the unconditional density. Hence, Lemma 7 yields $\mathbf{Pr}\left[\Lambda < \varepsilon | \neg\mathcal{F}\right] \le \varepsilon \cdot 32cn^5m^7m_{\max}\phi^2p^2$. Furthermore, it holds $\mathbf{Pr}\left[\mathcal{F}\right] = \mathbf{Pr}\left[\mathcal{E} \cup \mathcal{G}\right] \le \mathbf{Pr}\left[\mathcal{E}\right] + \mathbf{Pr}\left[\mathcal{G}\right] \le \frac{1}{p}$. Thus, we obtain

$$\mathbf{Pr}\left[\Lambda < \varepsilon\right] \le \frac{1}{p} + \varepsilon \cdot 32cn^5m^7m_{\max}\phi^2p^2.$$

Setting $p = (\varepsilon \cdot 32cn^5m^7m_{\max}\phi^2)^{-1/3}$ yields the desired result. The upper bound on ε is due to the assumption $p \ge 1$. The claim about the feasibility gap follows analogously. □

2.3. Loser and feasibility gap for multiple constraints.

Assume there are $k \ge 2$ constraints. W.l.o.g. these constraints are of the form $Ax \le b$ with $A \in \mathbb{R}^{k \times n}$ and $b \in \mathbb{R}^k$, and the set of points satisfying these constraints are $\mathcal{B}_1, \dots, \mathcal{B}_k$, respectively. We generalize the definition of feasibility and loser gap as follows. Given a set of solutions $\mathcal{S}$ and a ranking, the winner x^* is the highest ranked solution in $\mathcal{S} \cap \mathcal{B}_1, \dots, \cap \mathcal{B}_k$. The *feasibility gap for multiple constraints* is the minimal slack of x^* over all constraints, that is, $\Gamma = \min_{j \in [k]}\{(b - Ax)_j\}$, if x^* exists, and $\Gamma = \perp$ otherwise.

A solution in $\mathcal{S}$ is called loser if it has a higher rank than x^*. Observe that a loser only needs to be infeasible with respect to one of the k constraints. In particular, it is not true that the weight values of each loser are likely to be far away from the corresponding thresholds t_j, $j \in [k]$; not even if we consider only those constraints for which the respective loser is infeasible. Fortunately, however, we do not need such a property in the application of the loser gap. For every loser, one only needs a single constraint that renders the loser infeasible. Therefore, we define the *loser gap for k constraints* by

$$\Lambda = \begin{cases} \min_{x\in\mathcal{L}} \max_{j\in[k]}\{w_j^T x - t_j\} \text{ if } \mathcal{L} \neq \emptyset & \text{, and} \\ \perp & \text{otherwise.} \end{cases}$$

Lemma 8. *Let $c = \max_{j\in[k]} \max_{i\in[n]} \mathbf{E}\left[|A_{j,i}|\right]$. Then, for all ε with $\varepsilon \leq (32n^5m^7m_{\max}\phi^2)^{-1}$, $\mathbf{Pr}\left[\Gamma \leq \varepsilon\right] \leq 2k(\varepsilon \cdot 32cn^5m^7m_{\max}\phi^2)^{1/3}$ and $\mathbf{Pr}\left[\Lambda \leq \varepsilon\right] \leq 2k(\varepsilon \cdot 32cn^5m^7m_{\max}\phi^2)^{1/3}$.*

Proof. First we show the bound for the feasibility gap. Let x^* denote the winner and suppose $\Gamma \leq \varepsilon$, for some $\varepsilon \in \mathbb{R}_{\geq 0}$. Then there exists a $j \in [k]$ with $t_j - w_j^T x^* \leq \varepsilon$. Thus,

$$\mathbf{Pr}\left[\Gamma \leq \varepsilon\right] \leq \sum_{j\in[k]} \mathbf{Pr}\left[t_j - w_j^T x^* \leq \varepsilon\right] \quad .$$

For each individual $j \in [k]$, we can apply the Separating Lemma assuming that the set of feasible solutions with respect to all other constraints is fixed as the coefficients in this constraint are stochastically independent from the other constraints. This way, we obtain $\mathbf{Pr}\left[\Gamma \leq \varepsilon\right] \leq 2k(\varepsilon \cdot 32cn^5m^7m_{\max}\phi^2)^{1/3}$.

Next, we turn our attention to the loser gap. Unfortunately, we cannot generalize the bound on the loser gap from one to multiple constraints in the same way as we generalized the feasibility gap since the loser gap for multiple constraints does not correspond to the minimal loser gap over the individual constraints. Instead we will make use of the result for the feasibility gap established above. Assume $\Lambda \leq \varepsilon$, for some $\varepsilon \in \mathbb{R}_{\geq 0}$.

Then there exists a loser x satisfying $\forall j \in [k] : w_j^T x - t_j \leq \varepsilon$. Let x_L denote the loser with this property that is ranked highest. Consider a relaxed variant I' of the given ILP I where the thresholds of all stochastic constraints are increased by ε, i.e., we have constraints $w_j^T x \leq t_j + \varepsilon, j \in [k]$. Observe that x_L is feasible in the relaxed ILP I' and, by the definition of x_L, no higher ranked solution is feasible. Thus, x_L is the winner of I'. Since $t_j < w_j^T x_L \leq t_j + \varepsilon$ for some $j \in [k]$, the feasibility gap Γ' of the relaxed problem is smaller than ε. Hence, $\Lambda \leq \varepsilon$ implies $\Gamma' \leq \varepsilon$. Finally, applying the bound $\mathbf{Pr}[\Gamma' \leq \varepsilon] \leq 2k(\varepsilon \cdot 32cn^5m^7m_{\max}\phi^2)^{1/3}$ derived in the first part of the proof yields $\mathbf{Pr}[\Lambda \leq \varepsilon] \leq 2k(\varepsilon \cdot 32cn^5m^7m_{\max}\phi^2)^{1/3}$. □

3. From structural properties to algorithms

We will now describe how the isolating and the Separating Lemma, i.e. Lemma 1 and 2, can be used for turning an algorithm with pseudopolynomial worst-case complexity into an algorithm with polynomial smoothed complexity. In order to apply the pseudopolynomial algorithm efficiently it seems to be necessary to round the coefficients but obviously even the smallest rounding can change the optimal solution in a deterministically chosen ILP. However, using the results about the sizes of winner, loser and feasibility gaps shown in the previous section we are able to show that typically rounding the coefficients of a semi-random ILP after a logarithmic number of bits does not change the optimal solution.

We will exploit this property by an *adaptive rounding* approach. To be more precise, our algorithm will round the coefficients after a logarithmic number of bits d and call the pseudopolynomial time algorithm $\mathcal{A}$ to obtain an optimal solution x' for the rounded input in polynomial time. After that, the algorithm will call a *certifier*, i.e. a subroutine determining whether x' is also the optimal solution of the original, non-rounded, ILP or not. In the former case the algorithm can stop with output x', in the latter case the precision d is increased by one and the algorithm starts over again.

In the following sections we will first present and analyze the certifiers and, after that, we will show that the adaptive rounding approach yields polynomial smoothed running time.

3.1. A certifier for stochastic objective functions

First we will consider the case that only the coefficients in the objective function are randomly perturbed. The certifier used in [5] heavily relies on the fact that the variables are binary. Hence, we have to substantially change it in order to make it work in the general case of ILPs.

The crucial observation is that rounding after the d-th bit after the binary point changes each coefficient in the objective function by at most 2^{-d}. Since the absolute value of every variable is bounded by $m_{\max}$ we conclude that rounding changes the objective value of each solution $x \in \mathcal{S}$ by at most $nm_{\max}2^{-d}$. Thus, if the optimal solution x' of the rounded ILP is better by at least $2nm_{\max}2^{-d}$ than every other solution then we know for sure that x' is also the optimal solution of the original, non-rounded, ILP.

For any real number (or matrix of real numbers) a, let $\lfloor a \rfloor_d$ denote the number (or matrix) obtained from a by rounding it (or all its components) down after the d-th bit after the binary point, i.e. by cutting of the remaining bits. Furthermore, let x' denote the optimal solution of the rounded ILP, i.e. x' is the optimal solution of $\max(\lfloor c \rfloor_d)^T x$ w.r.t. the constraints $Ax \leq b$.

The certifier calls the pseudopolynomial time algorithm $2n$ many times, each time with slightly modified coefficients in the objective function. Let $c^{(i)}$ denote the coefficients in the i-th call of the pseudopolynomial time algorithm $\mathcal{A}$ and let $x^{(i)}$ be the optimal solution to the objective function $\max(c^{(i)})^T x$ w.r.t. the constraints $Ax \leq b$. For $i \in \{1, \ldots, n\}$ we define $c^{(i)}$ to be

$$c_j^{(i)} = \begin{cases} \lfloor c_j \rfloor_d & \text{if } j \neq i \\ \lfloor c_j \rfloor_d + (nm_{\max} + 1)2^{-d+1} & \text{if } j = i \end{cases}$$

and for $i \in \{n+1, \ldots, 2n\}$ we define $c^{(i)}$ to be

$$c_j^{(i)} = \begin{cases} \lfloor c_j \rfloor_d & \text{if } j \neq i - n \\ \lfloor c_j \rfloor_d - (nm_{\max} + 1)2^{-d+1} & \text{if } j = i - n \end{cases}.$$

That is, in each of the first n rounds one of the coefficients is slightly increased and in each of the last n rounds one of the coefficients is slightly decreased. If, for all $i \in [2n]$, $x^{(i)} = x'$, i.e. x' is the optimal solution to all of these ILPs, then we know for sure that x' is also the optimal solution to the non-rounded objective function $c^T x$, otherwise we cannot certify x' to be the true winner and report a failure. In the latter case, which can only occur if the winner gap is small, the adaptive rounding algorithm will increase the precision d and start over again. Now we will prove these properties formally.

Lemma 9. *If $x' = x^{(i)}$ for all $i \in [2n]$ then x' is the true winner, i.e. the optimal solution of the non-rounded ILP. If there exists an $i \in [2n]$ with $x' \neq x^{(i)}$ then the winner gap cannot be larger than $nm_{\max}2^{-d+1}$.*

Proof. Assume $x' = x^{(i)}$ for all $i \in [2n]$ and assume there exists a solution $y \in \mathcal{S}$ with $Ay \le b$ and $c^T y > c^T x'$, that is, y is a better solution of the non-rounded ILP than x'. Since $x' \ne y$, we can choose an index $i \in [n]$ with $x'_i \ne y_i$. First, consider the case $x'_i < y_i$. We will show that in this case we have $x^{(i)} = y$, which contradicts the assumption $x^{(i)} = x'$.

Because $c^T y > c^T x' = c^T x^{(i)}$ and since rounding changes the objective values of y and $x^{(i)}$ by at most $nm_{\max}2^{-d}$ we have

$$(\lfloor c \rfloor_d)^T y - (\lfloor c \rfloor_d)^T x' \ge -nm_{\max}2^{-d+1}$$

and thus

$$\begin{aligned}
&(c^{(i)})^T y - (c^{(i)})^T x' \\
&= (\lfloor c \rfloor_d)^T y - (\lfloor c \rfloor_d)^T x' + (y_i - x'_i)(nm_{\max} + 1)2^{-d+1} \\
&\ge (\lfloor c \rfloor_d)^T y - (\lfloor c \rfloor_d)^T x' + (nm_{\max} + 1)2^{-d+1} \\
&\ge -nm_{\max}2^{-d+1} + (nm_{\max} + 1)2^{-d+1} \\
&= 2^{-d+1} > 0.
\end{aligned}$$

Hence, y is a better solution to the objective function $(c^{(i)})^T x$ than x' which contradicts the assumption $x' = x^{(i)}$. Thus, if $x' = x^{(i)}$ for every $i \in [2n]$, then x' must be the optimal solution of the non-rounded ILP. In the case $x'_i > y_i$ one can analogously show that y is a better solution to the objective function $(c^{(n+i)})^T x$ than x'.

Now assume there exists an $i \in [2n]$ with $x' \ne x^{(i)}$. There are two cases between we have to distinguish: either x' is the true optimum of the non-rounded ILP or x' has become the optimal solution only due to the rounding. The case that x' is not the true optimal solution of the non-rounded ILP can occur only if the winner gap is not larger than $nm_{\max}2^{-d+1}$ since

$$(\lfloor c \rfloor_d)^T x^* - (\lfloor c \rfloor_d)^T x' \geq c^T x^* - c^T x' - nm_{\max}2^{-d+1} \geq \Delta - nm_{\max}2^{-d+1}.$$

Now assume x' to be the true optimum of the non-rounded ILP, i.e. $x' = x^*$, and $x' \neq x^{(i)}$ for an $i \in [2n]$, that is $(c^{(i)})^T x^* \leq (c^{(i)})^T x^{(i)}$. Then we have

$$\Delta \leq c^T x^* - c^T x^{(i)} \leq (c^{(i)})^T x^* - (c^{(i)})^T x^{(i)} + nm_{\max}2^{-d+1} \leq nm_{\max}2^{-d+1}.$$

Thus, also in this case the winner gap cannot exceed $nm_{\max}2^{-d+1}$. □

Corollary 2. *The probability that the certifier fails since more bits need to be revealed to determine the true optimal solution is bounded from above by* $n^2 m^2 m_{\max} \phi 2^{-d+1}$.

Proof. If the certifier fails the winner gap cannot be larger than $nm_{\max}2^{-d+1}$. Applying Lemma 1 yields the desired bound. □

3.2. A certifier for stochastic constraints

Let I denote a semi-randomly created ILP and let k denote the number of constraints, that is the constraints have the form $Ax \leq b$ with $A \in \mathbb{R}^{k \times n}$ and $b \in \mathbb{R}^k$. Furthermore, let x' denote the optimal solution of the rounded ILP $\max c^T x$ w.r.t. $A'x \leq b' + (nm_{\max} + 1)2^{-d}$ with $A' = \lfloor A \rfloor_d$ and $b' = \lfloor b \rfloor_d$. In the following we assume that there is a unique ranking among the solutions, that is, we assume there are no two solutions with the same objective values. This is without loss of generality since we can assume solutions with equal objective values to be ranked in an arbitrary order.

The certifier consists of only two steps: First, it tests whether there exists an index $i \in [k]$ such that $b_i \in [-(nm_{\max}+1)2^{-d}, 0)$ or not. In the former case, x' cannot be certified to be the true winner with the given number of revealed bits d per coefficient. In the latter case the certifier tests whether $A'x' \leq b' - (nm_{\max}+1)2^{-d}$ holds or not. In the former case x' can be certified to be the true winner, i.e. the optimal solution of the non-rounded ILP I. Otherwise x' cannot be certified to be the true winner. The latter case can only occur if either the loser or the feasibility gap is small.

Lemma 10. *If $A'x' \leq b' - (nm_{\max}+1)2^{-d}$ then x' is the optimal solution of the non-rounded ILP I. If $A'x' \not\leq b' - (nm_{\max}+1)2^{-d}$ then either the loser gap or the feasibility gap does not exceed $nm_{\max}2^{-d+1}$.*

Proof. First, we will show that solutions feasible w.r.t. the constraints $Ax \leq b$ stay feasible w.r.t. the constraints $A'x \leq b' + (nm_{\max}+1)2^{-d}$. Assume $Ax \leq b$. Since the rounding changes each coefficient by at most 2^{-d} and since, for $i \in [n]$, it holds $|x_i| \leq m_{\max}$, the *j-th weight* of the solution x, i.e. $(Ax)_j := a_{j,1}x_1 + \ldots + a_{j,n}x_n$, for $j \in [k]$, is changed by at most $nm_{\max}2^{-d}$. Hence,

$$A'x = \lfloor A \rfloor_d x \leq Ax + nm_{\max}2^{-d} \leq b + nm_{\max}2^{-d} \leq \lfloor b \rfloor_d + (nm_{\max}+1)2^{-d},$$

where $\leq$ means $\leq$ in every component and, for a matrix A and a real number z, $A + z$ denotes the matrix obtained from A by adding z to each entry. Thus, there cannot be a feasible solution of the ILP I which is ranked higher than x'.

Now we have to make sure, that x' is a feasible solution of the non-rounded ILP and has not become feasible only due to the rounding. If $A'x' \leq b' - (nm_{\max}+1)2^{-d}$

then x' must also be a feasible solution of I, i.e. $Ax \leq b$, since

$$\begin{aligned} Ax' &\leq A'x' + nm_{\max}2^{-d} \\ &\leq (b'x' - (nm_{\max}+1)2^{-d}) + nm_{\max}2^{-d} \\ &= b'x' - 2^{-d} \leq bx'. \end{aligned}$$

Hence, in this case, the rank of x' cannot be higher than the rank of the true winner x^*. Altogether, the rank of x' must equal the rank of x^* and thus x' must equal x^*.

Now we have to consider the case that $A'x' \leq b' - (nm_{\max}+1)2^{-d}$ does not hold and show that, in this case, either the loser of the feasibility gap has to be small. We distinguish between the case that x' is the true optimal solution of the non-rounded ILP I, that is $x' = x^*$, and the case that x' has become feasible only due to the rounding. In the former case we have $Ax' \leq b$ and there exists an $l \in [k]$ with $b'_l - (A'x')_l < (nm_{\max}+1)2^{-d}$. Hence, we have

$$\begin{aligned} \Gamma &\leq \min_{j\in[k]} \{b_j - (Ax')_j\} \\ &\leq b_l - (Ax')_l \\ &\leq b'_l - (A'x')_l + (nm_{\max}+1)2^{-d} \\ &\leq (nm_{\max}+1)2^{-d+1}, \end{aligned}$$

that is, the feasibility gap cannot be larger than $(nm_{\max}+1)2^{-d+1}$.

In the latter case we have $(A'x')_l \leq b'_l + (nm_{\max}+1)2^{-d}$ and hence

$$\begin{aligned} \Lambda &\leq \max_{j\in[k]} \{(Ax')_j - b_j\} \\ &\leq \max_{j\in[k]} \left\{(A'x')_j - b'_j + (nm_{\max}+1)2^{-d}\right\} \\ &\leq (nm_{\max}+1)2^{-d+1}, \end{aligned}$$

that is, the loser gap cannot be larger than $(nm_{\max}+1)2^{-d+1}$. $\square$

Lemma 11. *Assume* $c = \max_{j\in[k]} \max_{i\in[n]} \mathbf{E}\left[|A_{j,i}|\right]$ *to be constant. The probability that the certifier fails is bounded from above by* $2^{-d} \cdot \mathrm{poly}(n, m, m_{\max}, \phi, k)$.

Proof. The certifier can only fail if, for at least one $i \in [k]$, $b_i \in [-(nm_{\max}+1)2^{-d}, 0)$ holds or if either the loser or the feasibility gap does not exceed $(nm_{\max}+1)2^{-d+1}$. Since the thresholds are random variables whose densities are bounded by ϕ, the probability of the first event is bounded from above by $k(nm_{\max}+1)2^{-d}\phi$. But observe that we cannot use Lemma 2 directly to bound the probability that one of the gaps is too small since this lemma is true only if 0^n is not a feasible solution.

First, assume that there exists at least one $j \in [k]$ such that $b_j < -(nm_{\max}+1)2^{-d}$. Then 0^n is not a feasible solution w.r.t. $Ax \le b$ and $A'x \le b' + (nm_{\max}+1)2^{-d}$ and hence, it does not affect the certifier. Thus, in this case, we can use the separating lemma directly to bound the probability that the certifier fails. Since the certifier fails if, for at least one $j \in [k]$, $b_j \in [-(nm_{\max}+1)2^{-d}, 0)$, in this case, the solution 0^n does not affect the certifier either.

Now assume $b_j \ge 0$ for all $j \in [k]$. Then 0^n is feasible w.r.t. $Ax \le b$ and $A'x \le b' + (nm_{\max}+1)2^{-d}$. If 0^n is not the optimal solution w.r.t. $Ax \le b$, then 0^n does not affect loser and feasibility gap. Hence, the only case which needs to be considered in more detail is the case that 0^n is the optimal solution w.r.t. $Ax \le b$. Observe that the feasibility of the solution 0^n can be verified easily. Therefore, no problem occurs in the case that 0^n is the optimal solution w.r.t. $A'x \le b' + (nm_{\max}+1)2^{-d}$.

The only case which is a little bit tricky to handle is the case that 0^n is the optimal solution w.r.t. $Ax \le b$ but that $x' \ne 0^n$ is the optimal solution w.r.t. $A'x \le b' +$

$(nm_{\max}+1)2^{-d}$. In this case, x' is rejected by the certifier since $\lfloor A \rfloor_d x' \leq \lfloor b \rfloor_d - (nm_{\max}+1)2^{-d}$ does not hold. We have to bound the probability that this case occurs. Analogously to the case $0^n \notin \mathcal{S}$, one can argue that this can only happen if the size of the loser gap Λ does not exceed $(nm_{\max}+1)2^{-d+1}$. Unfortunately, we cannot apply Lemma 8 directly since we analyzed the gaps only in the case $0^n \notin \mathcal{S}$. Instead, we exclude 0^n from the set of feasible solutions, that is we define $\mathcal{S}' = \mathcal{S} \backslash \{0^n\}$ and argue with the help of the loser gap Λ' w.r.t. $\mathcal{S}'$. The crucial observation is that adding 0^n to the set of solutions can, in the case $b \geq 0$, only result in an increase of the size of the loser gap. The reason therefore is that, in the case $b \geq 0$, 0^n is a feasible solution which means that by adding 0^n to the set of solutions one cannot enlarge the set of losers $\mathcal{L}$. Hence, it holds $\Lambda \geq \Lambda'$ and we can make use of Lemma 8 in order to bound the probability that Λ' does not exceed $(nm_{\max}+1)2^{-d+1}$.

Altogether, the failure probability is bounded by

$$\begin{aligned} & \mathbf{Pr}\left[\exists i \in [k] : b_i \in [-(nm_{\max}+1)2^{-d}, 0)\right] \\ & + \mathbf{Pr}\left[\Gamma \leq (nm_{\max}+1)2^{-d+1}\right] + \mathbf{Pr}\left[\Lambda \leq (nm_{\max}+1)2^{-d+1}\right] \\ \leq \quad & 2^{-d} \cdot \mathrm{poly}(n, m, m_{\max}, \phi, k). \end{aligned}$$

□

If $\mathcal{D} \subset \mathbb{N}_0$ then it is not necessary to increase the threshold b' by $(nm_{\max}+1)2^{-d}$ for calculating x'. We only increased b' to ensure that feasible solutions of $Ax \leq b$ stay feasible after the rounding. If $x \geq 0$ and we round down each entry of A then this is already ensured by the rounding. Observe, that in this case no problem with the solution 0^n occurs since it is either feasible w.r.t. both $Ax \leq b$ and $A'x \leq b'$ or it is infeasible w.r.t. both. Thus, if $\mathcal{D} \subset \mathbb{N}_0$ it is not important that none of the thresholds lies in the interval $[-(nm_{\max}+1)2^{-d}, 0)$, and hence, it is not necessary to perturb the thresholds.

3.3. A certifier for stochastic objective functions and constraints

If both the objective function and the constraints are perturbed, we can combine the certifiers described in the preceding sections. That is, first we round the coefficients in the constraints and use the certifier for stochastic objective functions to calculate a certified optimum x' w.r.t. the rounded constraints. Then we use the certifier for stochastic constraints to test whether x' is also an optimal solution of the non-rounded ILP. The combined certifier fails if one of the two certifier used as subroutines fails, hence the probability of failure can be bounded as in Lemma 11.

3.4. Adaptive Rounding

In this section we will analyze the adaptive rounding approach formally. Assume that we start with $d = 1$, that is we reveal only the first bit after the binary point of each coefficient. The pseudopolynomial algorithm is called to calculate the optimum w.r.t. the rounded coefficients and, after that, an appropriate certifier is called. If this certifier fails, the number of revealed bits d is increased by one and the pseudopolynomial algorithm and the certifier are called again. This is repeated until a certified winner can be calculated.

Lemma 12. *The adaptive rounding algorithm has polynomial smoothed running time.*

Proof. We will show that a polynomial P with the same properties as the one in (1) exists. Consider a class Π of ILPs and let I be an ILP from Π with n integer variables and

k stochastic expressions. Furthermore, let N denote the input length of the ILP I. Since each stochastic coefficient has a virtual length of 1, it holds $N \geq nk$ and a random perturbation does not change the length of I. Let $\mathcal{A}$ denote the (possibly randomized) pseudopolynomial algorithm and let $T(I)$ denote a random variable describing the running time of algorithm $\mathcal{A}$ on Input I plus the running time of the subsequently called certifier.

If d bits of each stochastic coefficient and threshold are revealed, then we scale each stochastic expression by the factor 2^d to obtain integral expressions. Let I_d denote the rounded and scaled ILP obtained this way and observe that each stochastic number in I_d is rounded and scaled in such a way that it is integral. Let W denote the largest absolute value taken by one of the stochastic numbers in I_d, then $W = W_1 W_2$, where the factor $W_1 = 2^d$ is due to the scaling and the factor W_2 denotes the integer part of the largest absolute value of any stochastic number before the scaling. Since the running time of $\mathcal{A}$ is pseudopolynomially bounded (w.r.t. the perturbed numbers), we can choose two constants $c_1, c_2 \in \mathbb{R}$ with $c_2 \geq 1$ such that for each ILP I from the class Π with length N we have $\mathbf{E}\left[T(I_d)\right] \leq c_1(NW)^{c_2} = c_1(N2^d W_2)^{c_2}$.

To analyze the running time of this adaptive rounding approach we have to estimate $W_1 = 2^d$ and W_2. If the certifier concludes optimality after d_0 bits after the binary point of each random number have been revealed, we obtain the following estimate on the expected running time $\mathbf{E}[T_{\mathrm{AR}}]$ of the adaptive rounding

$$\begin{aligned}\mathbf{E}[T_{\mathrm{AR}}] &= \sum_{d=1}^{d_0}(\mathbf{E}[T(I_d)] + cN) \\ &\leq \sum_{d=1}^{d_0}(c_1(N2^dW_2)^{c_2} + cN) \\ &\leq cd_0N + c_1(N2^{d_0+1}W_2)^{c_2} \qquad (8)\end{aligned}$$

where cN denotes the costs for revealing an additional bit of each random number and for scaling the constraints. Hence, we have to estimate how large the values of d_0 and W_2 are typically. Since the absolute mean value of a random variable which is described by the density f_ϕ is bounded by $E/\phi \leq E$ for some constant $E \in \mathbb{R}$ and since we assume the stochastic numbers to be in the interval $[-1, 1]$ before perturbing them, an easy application of Markov's inequality and a union bound show $\mathbf{Pr}[W_2 > (E+1) \cdot 4N/\varepsilon] \leq \varepsilon/4$.

We have seen (Corollary 2 and Lemma 11) that, for an appropriate polynomial p, the probability that the certifier fails after d bits after the binary point of each coefficient have been revealed can be bounded by $2^{-d} \cdot p(n, m, m_{\max}, \phi, k)$. Hence, there is a polynomial q with $\mathbf{Pr}[d_0 > \log(q(N, \phi, 1/\varepsilon))] \leq \varepsilon/2$.

In equation (8), we substitute d_0 by $\log(q(N, \phi, 1/\varepsilon))$, W_2 by $(E+1) \cdot 4N/\varepsilon$ and multiply the resulting polynomial by $4/\varepsilon$. We denote the polynomial obtained this way by P. For all $N \in \mathbb{N}$, $\phi \geq 1$ $\varepsilon \in (0,1]$ and for all $I \in I_N$ we have

$$\begin{aligned}
&\mathbf{Pr}\left[T_{\mathrm{AR}}(I+f_\phi) \geq P\left(N,\phi,\frac{1}{\varepsilon}\right)\right] \\
\leq\; &\mathbf{Pr}\left[d_0 > \log(q(N,\phi,1/\varepsilon))\right] + \mathbf{Pr}\left[W_2 > (E+1)\cdot 4N/\varepsilon\right] \\
&+\mathbf{Pr}\left[T_{\mathrm{AR}}(I+f_\phi) > \frac{4}{\varepsilon}\mathbf{E}\left[T_{\mathrm{AR}}(I+f_\phi)\right]\right] \\
\leq\; &\frac{\varepsilon}{2}+\frac{\varepsilon}{4}+\frac{\varepsilon}{4} = \varepsilon,
\end{aligned}$$

since the running time of the adaptive rounding can exceed the bound given by the polynomial P only if either d_0 or W_2 exceed their bounds or if the expected running time of the adaptive rounding is larger by a factor of at least $4/\varepsilon$ than its expected running time. □

We have proven, how an algorithm with pseudopolynomial worst-case complexity can be turned into an algorithm with polynomial smoothed complexity. This concludes the first part of the proof of Theorem 1.

4. From polynomial smoothed complexity to pseudopolynomial running time.

Finally, we need to show that polynomial smoothed complexity implies the existence of a randomized pseudopolynomial algorithm. This can be shown by arguments similar to those used in the analysis of the binary case [5]. For the sake of completeness we provide the proof here.

Proof (Second part of Theorem 1).

Since we are aiming for a pseudo-polynomial time algorithm, we can assume that all numbers in the stochastic expressions are integers. Let M denote the largest absolute value of these numbers. The idea is to perturb all numbers only slightly such that the perturbation changes the value of each expression by at most $\frac{1}{2}$. To ensure that the set of feasible solutions is not changed by the perturbation, we relax all constraints by $\frac{1}{2}$, i.e., we replace $w^T x \leq t$ by $w^T x \leq t + \frac{1}{2}$ for all stochastic constraints. We then use an algorithm with polynomial smoothed complexity to compute an optimal solution x^* for the perturbed problem. By bounding the error due to the random perturbation, x^* can be shown to be optimal for the original problem as well.

Let us describe the proof in more detail. Our smoothed analysis framework assumes that all numbers in the stochastic expressions fall into the interval $[-1, 1]$ (or $[0, 1]$) before they are perturbed. To adapt our problem to this framework, we first scale all input numbers in the stochastic expressions by M^{-1} and adapt the thresholds accordingly, i.e., $w_1 x_1 + w_2 x_2 + \ldots + w_n x_n \leq t$ is replaced by $(w_1/M)x_1 + (w_2/M)x_2 + \ldots + (w_n/M)x_n \leq t/M$. Consequently, we have to ensure that the perturbation changes the value of an expression by at most $1/(2M)$. In particular, we will allow only perturbations that change each individual number by at most $1/(2Mnm_{\max})$. We call such a perturbation *proper*. For the uniform distribution, we could simply set $\phi = 2Mnm_{\max}$. However, we have to deal with arbitrary families of distributions, as defined in our smoothed analysis framework, and they do not necessarily have a finite domain. The idea is to choose ϕ large enough so that a random perturbation is proper with probability at least $1/2$. Recall that the perturbation model is described by the density function f with density parameter $\phi = 1$. For other values of ϕ, we scale f appropriately. By our assumptions on f, it holds $\int |t| f_\phi(t) dt = E/\phi$ for some fixed $E \in \mathbb{R}$. Let r be a random variable following f_ϕ. Setting $\phi = 4n^2 m_{\max} kEM$ and applying Markov's inequality yields $\mathbf{Pr}\left[|r| > \frac{1}{2nm_{\max}M}\right] = \mathbf{Pr}\left[|r| > \frac{2nkE}{\phi}\right] \leq \frac{1}{2nk}$. Our perturbation draws kn of these random variables. The probability that the perturbation is proper, i.e., the probability that their absolute values are at most $\frac{1}{2nm_{\max}M}$, is $1/2$.

Consider any binary optimization problem Π with polynomial smoothed complexity. Polynomial smoothed complexity implies that the problem admits an algorithm $\mathcal{A}$ whose running time can be bounded polynomially in n and ϕ, with arbitrary large constant probability strictly less than 1. In particular, there exists a polynomial $P(n, \phi)$ such that the probability that the running time exceeds $P(n, \phi)$ is at most $\frac{1}{4}$. We use $\mathcal{A}$ as a subroutine in order to obtain a pseudo-polynomial algorithm. This algorithm works as follows. At first, it generates a perturbation and checks whether it is *proper*. If it is proper, then it runs $\mathcal{A}$ for at most $P(n, \phi)$ time steps. If $\mathcal{A}$ has not finished within this time bound, the algorithm returns FAILURE. Let Q be the event that the perturbation is proper. Observe that for every two events A and B it holds $\mathbf{Pr}[A \wedge B] \geq \mathbf{Pr}[A] + \mathbf{Pr}[B] - 1$. Therefore, the success probability of our algorithm is

$$\mathbf{Pr}[Q \wedge (T \leq P(n, \phi))] \geq \mathbf{Pr}[Q] - \mathbf{Pr}[T > P(n, \phi)] \geq \frac{1}{4} .$$

The running time of this algorithm is pseudo-polynomial because $\phi = O(Mn^2 m_{\max} k)$. Hence, $\Pi_u \in \mathsf{ZPP}$. This completes the proof of Theorem 1. □

5. Packing and Covering ILPs

In this section, we prove Theorem 2 and Theorem 3. For the sake of simplicity we consider only the case that the thresholds are fixed arbitrarily instead of being randomly perturbed. This is without loss of generality since perturbing the thresholds is not necessary if $\mathcal{D} \subseteq \mathbb{N}_0$, as we have already mentioned in Section 3.2, and furthermore it only strengthens the adversary.

First, we look at ILPs containing at least one perturbed packing constraint.

Proof (Theorem 2). We have to show the existence of a polynomial P with the same properties as the one in (1). We will show that one can add the box constraints $x_i \leq m$, $1 \leq i \leq n$, for a suitable chosen $m = \text{poly}(N, \varepsilon^{-1})$ such that with probability at least $1 - \varepsilon/2$ these additional box constraints do not change the optimal solution of the ILP. The class of ILPs with the additional box constraints has polynomial smoothed complexity due to Theorem 1. Combining this with the additional failure probability of at most $\varepsilon/2$ implies the theorem. Now, we will analyze this formally.

Consider an ILP with at least one perturbed packing constraint $a_1, \dots, a_n \leq 1$ where $a_1, \dots, a_n$ are random variables with density bounded by ϕ and let $x \in \mathbb{N}_0^n$ be an arbitrary solution fulfilling this constraint. For $i \in [n]$, if $x_i > m$ then $a_i < 1/m$. Since

$$\mathbf{Pr}\left[\exists i \in [n] : a_i < \frac{1}{m}\right] = \mathbf{Pr}\left[\exists i \in [n] : a_i \in \left[0, \frac{1}{m}\right)\right] \leq \frac{n\phi}{m},$$

if we set m to $\lceil 2n\phi/\varepsilon \rceil$, the probability that restricting the feasible region to $\{0, \dots, m\}^n$ changes the optimal solution is bounded by $\varepsilon/2$.

We only look at the restricted problem with $\mathcal{S} = \{0, \dots, m\}^n$. We can apply Theorem 1 to show that this problem has polynomial smoothed complexity. That means, it exist a polynomial P' and an algorithms A whose running time T satisfies

$$\mathbf{Pr}\left[T(I + f_\phi) \geq P'\left(N, \phi, \frac{1}{\varepsilon}\right)\right] \leq \frac{\varepsilon}{2},$$

for every $N \in \mathbb{N}$, $\phi \geq 1$, $\varepsilon \in (0, 1]$, $I \in \mathcal{I}_N$. W.l.o.g. we can assume P' to be of the form $P'(N, \phi, \varepsilon^{-1}) = m_{\max}^{\alpha} \cdot p(N, \phi, \varepsilon^{-1})$ for some constant α and a polynomial p independent of $\mathcal{S}$. We define a polynomial P by

$$P(N, \phi, \varepsilon^{-1}) = \beta \cdot \left(\left\lceil \frac{2n\phi}{\varepsilon} \right\rceil \right)^{\alpha} \cdot p(N, \phi, \varepsilon^{-1})$$

for some constant β to be defined later. Let $\mathcal{K}(\varepsilon)$ denote the event that there exists an $i \in [n]$ with $a_i < (\lceil 2n\phi/\varepsilon \rceil)^{-1}$, then, we have

$$\begin{aligned}
&\mathbf{Pr}\left[T(I + f_\phi) \geq P\left(N, \phi, \frac{1}{\varepsilon}\right) \right] \\
&\leq \mathbf{Pr}\left[\mathcal{K}(\varepsilon)\right] + \mathbf{Pr}\left[T(I + f_\phi) \geq P\left(N, \phi, \frac{1}{\varepsilon}\right) \middle| \neg\mathcal{K}(\varepsilon) \right] \\
&\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,
\end{aligned}$$

where the last inequality follows since under the condition $\neg\mathcal{K}(\varepsilon)$ the feasible region $\mathcal{S}$ is $\{0, \ldots, m\}^n$ with $m = \lceil 2nk\phi/\varepsilon \rceil$. Observe that the additional constant β is needed since the condition $\neg\mathcal{K}(\varepsilon)$ can increase the maximal density ϕ. But not more than by a factor of 2 since $\mathbf{Pr}\left[\neg\mathcal{K}(\varepsilon)\right] \leq 1/2$. $\square$

The line of arguments used to prove Theorem 2 can also be used to prove Theorem 3. It only remains to show that for an ILP consisting only of perturbed covering constraints one can choose an $m = \mathrm{poly}(N, \varepsilon^{-1})$ such that adding the box constraints $x_i \leq m$, for $i \in [n]$, does not change the optimal solution, with high probability.

Proof (Theorem 3). Consider an ILP consisting of k covering constraints $Ax \geq 1$ and let x^* denote an optimal solution of this ILP. Since the objective function has to be minimized and all coefficients are non-negative we can assume that each vector y with $y < x^*$ is infeasible. Hence $x_i^* > m$ implies that there must be a $j \in [k]$ with $a_{j,i} <$

$1/(m-1)$. Since

$$\mathbf{Pr}\left[\exists i \in [n], j \in [k] : a_{j,i} < \frac{1}{m-1}\right] \leq \frac{nk\phi}{m-1},$$

if we set m to $\lceil 2nk\phi/\varepsilon + 1\rceil$, the probability that restricting the feasible region to $\{0,\ldots,m\}^n$ changes the optimal solution is bounded by $\varepsilon/2$.

The remaining part of the proof is completely analogous to the proof of Theorem 2.

□

6. Extensions

In this section, we will briefly discuss two extensions to the class of problems and the class of perturbation models to which our analysis can be applied. That is, we will show that one does not need a linear objective function if only the constraints are perturbed and that coefficients set to zero by the adversary do not need to be perturbed. For the sake of simplicity we have not analyzed this in the previous sections.

6.1. Nonlinear Objective functions and constraints

If only the constraints are perturbed and the objective function is left untouched by the randomization then the objective function does not need to be linear. In Section 2.2 and in Section 2.3, where we analyzed loser and feasibility gap, we did not make use of the fact that the objective function is linear. Instead, we only assumed a fixed ranking among the solutions and hence, Lemma 2 and Lemma 8 are true for arbitrary objective functions. Also the certifier described in Section 3.2 does not make use of the fact that the objective function was assumed to be linear. That is, also the certifier works for arbitrary rankings and hence for arbitrary objective functions.

For many problems it does not make sense to perturb every constraint since some of the constraints might describe an underlying problem structure which should not be touched by the randomization. In the analysis of loser and feasibility gap, we assumed that the feasible region is given by the intersection of an arbitrary set $\mathcal{S} \subseteq \mathcal{D}^n$ with the half-spaces $\mathcal{B}_1, \ldots, \mathcal{B}_k$ determined by the linear constraints. Observe that one can explicitly distinguish between adversarial and stochastic constraints by encoding the adversarial constraints into the set $\mathcal{S}$ which is not affected by the randomization. This way, it is even possible to have non-linear adversarial constraints.

6.2. Zero-Preserving Perturbations

One criticism of the smoothed analysis of the Simplex algorithm is that the additive perturbations destroy the zero-structure of an optimization problem, as they replace zeros with small values. See also the discussion in [14]. The same criticism applies to the zero-structure of ILPs. It turns out, however, that our probabilistic analysis in Section 2 is robust enough to allow to preserve zero-entries. In particular, we can extend our semi-random input model introduced in Section 1.1 by allowing the coefficients in the stochastic expressions to be fixed to zero instead of being a random variable. In the model of smoothed analysis, this corresponds to strengthening the adversary by avoiding the perturbation of these zero-coefficients.

Consider the expression $w^T x$ and let Z be the set of indices i with w_i fixed to zero. We call two solutions $x, x' \in \mathcal{S} \subseteq \{0,1\}^n$ equivalent, if they differ only in positions contained in Z, e.g., if $x_i \neq x'_i \Rightarrow i \in Z$ holds. This way, Z defines equivalence classes on $\mathcal{S}$ with respect to the expression $w^T x$. Clearly, $w^T x$ evaluates to the same value for all solutions within the same equivalence class.

First we consider the Separating Lemma. Observe that only the highest ranked solution in each equivalence class is relevant for the loser and feasibility gap. This is because the winner and the minimal loser are Pareto optimal solutions. As all solutions within an equivalence class have the same weight, only the highest ranked solution of this class can become Pareto optimal. For the purpose of analysis, we can virtually remove all solutions from $\mathcal{S}$ that are not ranked highest within its equivalence class. This way, we can ignore variables x_i with $i \in Z$ and apply the Separation Lemma as before.

A similar argument can be used to show that the Generalized Isolating Lemma stays valid with respect to equivalence classes, that is, the winner gap is defined to be the difference in objective value between the best and the second best equivalence class. However, it might be very likely that there are many optimal solutions, as the winning equivalence class might have many solutions. We adapt the certifier described in Section 3.1 to this situation by making two changes. Let x' denote the solution of the rounded ILP as defined in Section 3.1 and assume w.l.o.g. that $c_{n-l+1}, \ldots, c_n$ are those coefficients set to zero. Then only the $x^{(i)}$ for $i \in \{1, \ldots, n-l\} \cup \{n+1, \ldots, 2n-l\}$ are considered and it is not checked whether $x' = x^{(i)}$ but if x' and $x^{(i)}$ are in the same equivalence class. After these modifications, the arguments in Lemma 9 also hold for the case of zero-preserving perturbations.

7. Conclusions

Our probabilistic analysis shows that important classes of ILPs with a fixed number of constraints have polynomial smoothed complexity. This means that random or randomly perturbed instances of such ILPs can be solved in polynomial time. The presented algorithmic framework giving these results uses algorithms with pseudopolynomial worst-case complexity as subroutines. Usually these pseudopolynomial time algorithms are based on dynamic programming. We want to remark that we do not believe that this approach is the most practical one to tackle ILPs of this kind. We expect that *branch and bound* and *branch and cut* heuristics are much faster than algorithms based on dynamic programming. The next challenging task is a smoothed analysis of these heuristics in order to theoretically explain their great success on practical applications. We think that the main contribution of this paper is to point out chances and limitations for such a probabilistic analysis.

Acknowledgements. This work was supported in part by the EU within the 6th Framework Programme under contract 001907 (DELIS) and by DFG grant Vo889/2-1.

References

1. C. Banderier, R. Beier, and K. Mehlhorn. Smoothed Analysis of Three Combinatorial Problems. In *Proc. 28th International Symposium on Mathematical Foundations of Computer Science (MFCS)*, volume 97, pages 198–207, 2003.
2. R. Beier and B. Vöcking. An Experimental Study of Random Knapsack Problems. In *Proc. of the 12th Annual European Symposium on Algorithms (ESA)*, pages 616–627, 2004.
3. R. Beier and B. Vöcking. Probabilistic Analysis of Knapsack Core Algorithms. In *Proc. of the 15th Annual Symposium on Discrete Algorithms (SODA)*, pages 468–477, 2004.

4. R. Beier and B. Vöcking. Random Knapsack in Expected Polynomial Time. In *Journal of Computer and System Sciences*, volume 69(3), pages 306–329, 2004.
5. R. Beier and B. Vöcking. Typical Properties of Winners and Losers in Discrete Optimization. In *Proc. of the 36th Annual ACM Symposium on Theory of Computing (STOC)*, pages 343–352, 2004.
6. K. H. Borgwardt and J. Brzank. Average Saving Effects in Enumerative Methods for Solving Knapsack Problems. In *Journal of Complexity*, volume 10, pages 129–141, 1994.
7. P. Crescenzi, V. Kann, M. Halldorsson, M. Karpinski, and G. Woeginger. A compendium of NP optimization problems. http://www.nada.kth.se/˜viggo/problemlist/compendium.html.
8. M. E. Dyer and A. M. Frieze. Probabilistic Analysis of the Multidimensional Knapsack Problem. In *Mathematics of Operations Research*, volume 14(1), pages 162–176, 1989.
9. M. Garey and D. Johnson. *Computers and Intractability*. Freeman, 1979.
10. A. Goldberg and A. Marchetti-Spaccamela. On Finding the Exact Solution to a Zero-One Knapsack Problem. In *Proc. of the 16th Annual ACM Symposium on Theory of Computing (STOC)*, pages 359–368, 1984.
11. G. S. Lueker. Average-Case Analysis of Off-Line and On-Line Knapsack Problems. In *Journal of Algorithms*, volume 19, pages 277–305, 1998.
12. B. Manthey and R. Reischuk. Smoothed Analysis of the Height of Binary Search Tress. Technical Report TR05-063, Electronic Colloquium on Computational Complexity, 2005.
13. H. Röglin and B. Vöcking. Smoothed Analysis of Integer Programming. In *Proc. of the 11th Conference on Integer Programming and Combinatorial Optimization (IPCO)*, pages 276–290, 2005.
14. D. A. Spielman and S.-H. Teng. Smoothed Analysis of Algorithms: Why The Simplex Algorithm Usually Takes Polynomial Time. In *Journal of the ACM*, volume 51(3), pages 385–463, 2004.