Smoothed Analysis of Multiobjective Optimization

Heiko Röglin*
Department of Quantitative Economics
Maastricht University
Maastricht, The Netherlands
Email: heiko@roeglin.org

Shang-Hua Teng†
Computer Science Department
University of Southern California
Los Angeles, CA
Email: shanghua.teng@gmail.com

Abstract— We prove that the number of Pareto-optimal solutions in any multiobjective binary optimization problem with a finite number of linear objective functions is polynomial in the model of smoothed analysis. This resolves a conjecture of René Beier [5]. Moreover, we give polynomial bounds on all finite moments of the number of Pareto-optimal solutions, which yields the first non-trivial concentration bound for this quantity. Using our new technique, we give a complete characterization of polynomial smoothed complexity for binary optimization problems, which strengthens an earlier result due to Beier and Vöcking [8].

Keywords-multiobjective optimization; Pareto-optimal solutions; smoothed analysis

1. Introduction

The decisions faced by economic entities are often very complex and involve several, usually conflicting, objectives. This has led to a tremendous amount of research in the area of *multiobjective optimization*, considering constrained optimization problems with several objective functions of the following form:

$$\textbf{minimize } w^1(x)\,, \ldots, \textbf{minimize } w^d(x)$$
$$\textbf{subject to } x \text{ in the feasible region } \mathcal{S}.$$

A well-established heuristic approach for dealing with such problems is to generate the set of *Pareto-optimal solutions*. These are the solutions that are not dominated by other solutions, that is to say, a solution is Pareto-optimal if there does not exist another solution that is simultaneously better in all criteria. In practice, often the set of Pareto-optimal solutions, or *Pareto set* for short, is generated in order to filter out unreasonable trade-offs. Then some possibly human-assisted post-processing is applied to make a choice among the Pareto-optimal solutions.

Clearly this approach is only feasible if few solutions are Pareto-optimal. Otherwise, generating the Pareto set is too costly and it provides not enough guidance to the decision-maker. In many applications, it has been observed that typically the Pareto set is indeed small. A theoretical explanation for this is, however, still lacking. This is mostly due to the fact that, in the worst case, almost every problem with more than one objective can have an exponentially large Pareto set. So in order to find a rigorous explanation for the practical observations, one has to deviate from the classical worst-case perspective.

That is why we study multiobjective optimization problems in the framework of *smoothed analysis*. This framework, originally introduced to explain the practical success of the simplex method [21], is based on a semi-random input model, in which an adversary specifies an arbitrary input that is subsequently slightly perturbed at random. This model can be viewed as a worst-case analysis with a less powerful adversary; the small amount of randomness rules out pathological worst-case instances that are rarely observed in practice but dominate the worst-case analysis. After its invention in 2001, smoothed analysis has been successfully applied in a variety of different contexts, e.g., to explain the practical success of local search methods [3], [13], heuristics for the knapsack problem [7], online algorithms [4], and clustering [2].

1.1. Multiobjective Binary Optimization & Beier's Conjecture

We consider a very broad class of multiobjective optimization problems, namely those that can be formulated as binary optimization problems with linear objective functions. To be precise, we study problems whose instances have the following form: there are an arbitrary set $\mathcal{S} \subseteq \{0,1\}^n$ of feasible solutions and d linear objective functions $w^i\colon \mathcal{S} \to \mathbb{R}$ of the form $w^i(x) = w^i_1 x_1 + \cdots + w^i_n x_n$, which are to be minimized over $\mathcal{S}$. As the set $\mathcal{S}$ can encode arbitrary combinatorial structures, our model covers a wide variety of optimization problems from mathematical programming, network design, and scheduling. For a given graph, the set $\mathcal{S}$ can encode, for example, (the incidence vectors of) all spanning trees, paths between two nodes, or Hamiltonian cycles. Hence, well-studied problems like the multiobjective spanning tree, shortest path, and traveling salesperson problem can naturally be formulated in our model.

As mentioned above, if the coefficients in the objective functions can be chosen by an adversary, then for almost

Most of this work was done while the authors were at Microsoft Research New England.

*Supported by a fellowship within the Postdoc-Program of the German Academic Exchange Service (DAAD).
†Supported by NSF grant CCR-0635102.

every problem with more than one objective, the size of the Pareto set is exponential. But what happens if the adversary is less powerful and the coefficients are subject to a small amount of random noise? Let us assume, for example, that independent Gaussians with mean 0 and small standard deviation are added to the adversarial coefficients. In practice this randomness can stem, for example, from measurement errors or numerical imprecision. For the bicriteria case, it has been shown that the expected number of Pareto-optimal solutions is polynomial in this smoothed model, no matter how the set $\mathcal{S}$ and the adversarial coefficients are chosen [7].

In his Ph.D. thesis, René Beier conjectures that this result extends to multiobjective optimization [5]: for any constant number of linear objective functions, the number of Pareto-optimal solutions is bounded by a polynomial in the model of smoothed analysis.

1.2. Our Contribution

In this paper, we resolve Beier's conjecture affirmatively. Our proof is based on a structural property of optimization problems that we dubbed *generalized loser gap* and that measures how many Pareto-optimal solutions can have similar objective values. Roughly speaking, we show that in semi-random instances of optimization problems, with high probability only a constant number of Pareto-optimal solutions can have objective values that are very close to each other. This implies that in expectation only a polynomial number of Pareto-optimal solutions fit into the hypercube $[-1, 1]^d$.

Our new technique shows not only that the expected number of Pareto-optimal solutions is polynomially bounded, but also that every constant moment is bounded by a polynomial. This is important because many heuristics for generating the Pareto set have (at least for more than two objectives) a running time that depends quadratically on the number of Pareto-optimal solutions (cf., for example, [18] for the multiobjective knapsack problem or [9] for the multiobjective shortest path problem). The bound on the constant moments implies that these algorithms have expected polynomial running time, which does not follow from a mere bound on the expected number. Additionally, it also gives the first non-trivial concentration bound on the number of Pareto-optimal solutions for bi- and multiobjective optimization problems, showing that deviations from the expected value are much more unlikely than a simple application of Markov's inequality would suggest.

Finally, using our new techniques, we give a complete characterization of *polynomial smoothed complexity* for the family of binary optimization problems. This strengthens an earlier result of Beier and Vöcking [8]. They show that a single-criterion binary optimization problem has polynomial smoothed complexity if and only if it can be solved in randomized pseudo-polynomial time in the worst case. It is, for example, well-known that the knapsack problem can be solved in time polynomial in the input size and the largest profit. Hence, Beier and Vöcking's characterization implies that it has polynomial smoothed complexity if the profits are perturbed. On the other hand, the traveling salesperson problem is strongly NP-hard, which implies according to the characterization that it does not have polynomial smoothed complexity if the edge lengths are perturbed, unless $\mathrm{ZPP} = \mathrm{NP}$.

However, Beier and Vöcking's characterization uses a weaker notion of polynomial smoothed complexity than originally proposed in [21]. Instead of showing that the expected running time in the smoothed model is polynomial, their theorem states only that with reasonable probability the running time is polynomial. We present a new algorithmic scheme to extend the characterization to the original notion of polynomial smoothed complexity. In particular, this implies that every problem that can be solved in randomized pseudo-polynomial time in the worst case can be solved in expected polynomial time on smoothed instances. This shows that, unlike in average-case complexity theory, the notion of expected polynomial running time is surprisingly robust against changes of the machine model in the model of smoothed analysis.

1.3. Related Work

There exists a vast body of literature that focuses on multiobjective optimization. In particular, many algorithms for generating the Pareto set of various optimization problems such as the (bounded) knapsack problem [15], [18], the multi-criteria shortest path problem [9], [14], [20], and the multi-criteria network flow problem [11], [17] have been proposed. Since for all these problems the number of Pareto-optimal solutions can be exponential (see, e.g., [12]), none of these algorithms runs in polynomial time in the worst case. In practice, however, generating the Pareto set is tractable in many situations. For instance, Müller-Hannemann and Weihe [16] study experimentally the number of Pareto-optimal solutions in multi-criteria shortest path problems. They consider examples that arise from computing the set of best train connections (in view of travel time, fare, and number of train changes) and conclude that in this application scenario generating the complete Pareto set is tractable even for large instances.

The special case of our model for two objective functions has been studied by Beier and Vöcking [7]. They show that the expected number of Pareto-optimal solutions is bounded from above by $O(n^4\phi)$ and from below by $\Omega(n^2)$, where ϕ is a parameter measuring the amount of randomness, which we will define formally in the next section. Later their analysis has been extended to integer optimization problems and the upper bound on the expected number of Pareto-optimal solutions has been improved to $O(n^2\phi)$ [6]. Ackermann et al. [1] consider bicriteria optimization problems with two semi-random objective functions. They show that it is

unlikely that there exist two Pareto-optimal solutions that have almost the same value in one of the objectives.

Another way of coping with Pareto sets of exponential size is to compute *approximate Pareto sets*. A solution x is *ε-dominated* by another solution x' if $w^i(x')/w^i(x) \leq 1+\varepsilon$ for all $i \in \{1,\dots,d\}$. A set $\mathcal{P}_\varepsilon$ is an *ε-approximation of a Pareto set* $\mathcal{P}$ if for any solution $x \in \mathcal{P}$, there is a solution $x' \in \mathcal{P}_\varepsilon$ that ε-dominates it. Papadimitriou and Yannakakis [19] show that for any Pareto set $\mathcal{P}$, there is an ε-approximation of $\mathcal{P}$ with polynomially (in the input size and $1/\varepsilon$) many points. They also give sufficient and necessary conditions for the existence of an FPTAS for computing approximate Pareto sets of multi-criteria optimization problems. Vassilvitskii and Yannakakis [22] and Diakonikolas and Yannakakis [10] show how to compute ε-approximate Pareto sets of small size.

In the following section, we will introduce the model that we study formally and in its full generality. Then, in Section 3, we will explain our techniques and introduce the notion of generalized loser gap. In Sections 4 and 5, we apply the generalized loser gap to bound the smoothed number of Pareto-optimal solutions and their moments. In the subsequent sections, we discuss some extensions and applications of our results.

2. Model and Notations

Let us begin with some notation: For $n \in \mathbb{N}$, we denote by $[n]$ the set $\{1,\dots,n\}$. For a vector $x \in \mathbb{R}^n$ and $i \in [n]$, we denote by x_i the i-th component of x. Furthermore for $I \subseteq [n]$, we denote by x_I the $|I|$-dimensional subvector obtained from x by removing all components that do not belong to I. We denote by $\log$ the logarithm to base 2.

We consider instances of d-dimensional combinatorial optimization problems that can be written in the following form: There is an arbitrary set $\mathcal{S} \subseteq Z^n$ of feasible solutions, where $Z \subseteq \mathbb{Z}$ denotes an arbitrary finite set of integers. That is, the feasible region is defined to be a set of n-dimensional integer vectors with entries from Z. For most of the paper, we consider the binary case $Z = \{0,1\}$ to keep the presentation simple. Only at the end in Section 6.3, we discuss extensions of our results to the general integer case. Furthermore, there are d objective functions $w^i\colon \mathcal{S} \to \mathbb{R}$ that associate d values with each feasible solution. For $i \in [d]$ and $x \in \mathcal{S}$, we will refer to $w^i(x)$ as the *i-th weight of solution x*, and for the sake of simplicity, we assume that all weights are to be minimized, even though this assumption is not important for our analysis. While w^d can be an arbitrary function that assigns a unique value to each solution in $\mathcal{S}$, we assume that $w^1,\dots,w^{d-1}$ are linear functions of the form $w^i(x) = w^i_1x_1 + \cdots + w^i_nx_n$. Slightly abusing notation, we will use w^i not only to refer to the function $w^i\colon \mathcal{S} \to \mathbb{R}$ but also to the vector $(w^i_1,\dots,w^i_n)$ of coefficients.

We say that a solution $x \in \mathcal{S}$ *dominates* another solution $y \in \mathcal{S}$ if, for every $i \in [d]$, $w^i(x) \leq w^i(y)$, and for one $i \in [d]$, $w^i(x) < w^i(y)$. A solution $x \in \mathcal{S}$ that is not dominated by any other solution is called *Pareto-optimal*, and we denote by $\mathcal{P} \subseteq \mathcal{S}$ the set of Pareto-optimal solutions.

In the semi-random input model that we study, the set $\mathcal{S}$ of feasible solutions and the last objective function w^d are chosen by an adversary. The coefficients w^i_j for $i \in [d-1]$ and $j \in [n]$ are random variables drawn independently according to densities $f^i_j\colon [-1,1] \to [0,\phi]$ for some parameter $\phi \geq 1$. Instead of determining these coefficients exactly, the adversary can only choose the density functions f^i_j. This gives him slightly more power than the two-step model in which he first chooses coefficients that are subsequently perturbed at random. Intuitively, the parameter ϕ measures the degree of randomness. The larger ϕ is, the more concentrated the distributions of the coefficients can be. For instance, the adversary can choose for each coefficient an interval of length $1/\phi$ in which it is uniformly distributed. This shows that if ϕ approaches infinity, the semi-random model approaches the worst-case model. On the other hand, if ϕ is constant, then our model is close to a uniform average-case analysis. We will state all our bounds in terms of the number of variables n and the maximal density ϕ.

The densities are defined on $[-1,1]$. This is not a severe restriction as all distributions with a bounded domain are covered by this model as they can be shifted and scaled. Notice, however, that scaling a larger domain to $[-1,1]$ increases the maximum density ϕ. Furthermore, our model also covers distributions with exponentially decaying tails like Gaussians and exponential distributions because such distributions take with probability exponentially close to one only values from a bounded domain of polynomial size. For Gaussian distributions, the maximal density ϕ is, for example, proportional to the reciprocal of the standard deviation σ.

We denote by q_d the number of Pareto-optimal solutions and our goal is to bound the expected value of q^c_d for $c \geq 1$ from above. We assume that the set $\mathcal{S}$, the objective function w^d, and the densities f^i_j are chosen by an adversary so as to maximize the expected value of q^c_d.

3. Generalized Loser Gap

For single-criterion optimization problems, Beier and Vöcking [8] introduced the notions of *winner gap* and *loser gap*. These are structural properties that measure how robust instances of optimization problems are against small changes of the coefficients in the objective function or in one of the constraints. Beier et al. [6] observe that the loser gap is closely related to the number of Pareto-optimal solutions and obtain, based on this observation, improved bounds for the expected number of Pareto-optimal solutions. In this section, we present a more general notion of loser gap that allows us to derive not only polynomial bounds for the expected

number of Pareto-optimal solutions but also for all constant moments of their distribution.

3.1. Loser Gap

In order to define the loser gap, let us consider a bicriteria optimization problem with an arbitrary set $\mathcal{S} \subseteq \{0,1\}^n$ of feasible solutions, an arbitrary weight function w^2, and a linear weight function w^1. We transform this into a single-criterion problem by putting a constraint on w^1, leading to the following problem: find a solution $x \in \mathcal{S}$ that minimizes the second weight $w^2(x)$ under the linear constraint $w^1(x) = w^1_1 x_1 + \cdots + w^1_n x_n \leq t$ for some threshold $t \in \mathbb{R}$. We will refer to the optimal solution to this problem as the *winner*, denoted by $x^\star$. Moreover, we call a solution $x \in \mathcal{S}$ a *loser* if it has a smaller second weight than the winner, that is, $w^2(x) < w^2(x^\star)$. By definition of the winner $x^\star$, all losers must have a first weight of more than t, that is, they are cut off by the linear constraint. We denote by $\mathcal{L}$ the set of all losers. If there does not exist a solution with a first weight of at most t, then $\mathcal{L}$ is defined to be $\mathcal{S}$. Now the *loser gap* $\Lambda(t)$ is defined to be the distance of the loser set from the threshold t, that is, $\Lambda(t) = \min\{w^1(x) \mid x \in \mathcal{L}\} - t$.

Intuitively, the loser gap measures how robust the instance is with respect to small changes of the coefficients $w^1_1, \ldots, w^1_n$. This is exploited in [8] in the following way: If, for $b \in \mathbb{N}$, we round down each coefficient w^1_i after the b-th bit after the binary point, then we obtain a new set of coefficients $\tilde{w}^1_i$ with $\tilde{w}^1_i \in [w^1_i - 2^b, w^1_i]$. Let $\tilde{x}^\star$ denote the solution $x \in \mathcal{S}$ that minimizes the second weight $w^2(x)$ under the constraint $\tilde{w}^1_1 x_1 + \cdots + \tilde{w}^1_n x_n \leq t$. As we round the coefficients down, the optimal solution $x^\star$ is also feasible with respect to the rounded coefficients. It can, however, happen that an initially unfeasible solution $x \in \mathcal{S}$ with $w^1(x) > t$ becomes feasible due to the rounding. But observe that as long as all losers in $\mathcal{L}$ stay infeasible, the optimal solution does not change, that is, if $\tilde{w}^1_1 x_1 + \cdots + \tilde{w}^1_n x_n > t$ for all $x \in \mathcal{L}$, then $x^\star = \tilde{x}^\star$. As every coefficient is changed by the rounding by at most 2^{-b}, the first weight of every solution is changed by at most $n2^{-b}$ by the rounding. Hence, if the loser gap $\Lambda(t)$ is larger than $n2^{-b}$, then rounding does not change the optimal solution. In that case, one can compute the optimal solution $x^\star$ by rounding all coefficients and solving the rounded instance, which, depending on b and the problem, can be much more efficient than solving the original unrounded instance directly.

Now what has this to do with Pareto-optimal solutions? The crucial observation is that the loser $\hat{x} = \operatorname{argmin}\{w^1(x) \mid x \in \mathcal{L}\}$ with the smallest first weight is a Pareto-optimal solution, and, even more important, there cannot be a Pareto-optimal solution with a first weight strictly between t and the weight $w^1(\hat{x})$ of the minimal loser. Assume for contradiction that there is a Pareto-optimal solution x with a first weight between t and $w^1(\hat{x})$. As x is not dominated by the winner $x^\star$, it must have a second weight smaller than $w^2(x^\star)$, which renders it a loser. However, this contradicts the choice of $\hat{x}$ as the loser with the smallest first weight. Hence, if the loser gap $\Lambda(t)$ is at least ε, then no Pareto-optimal solution can have a first weight in $(t, t+\varepsilon)$. Beier et al. [6] essentially prove that for every t and $\varepsilon \geq 0$, the probability of $\Lambda(t) \leq \varepsilon$ is bounded from above by $\varepsilon n \phi$, which shows that it is, for every fixed value t, unlikely to have a Pareto-optimal solution with a first weight close to t.

Beier et al. [6] count the Pareto-optimal solutions roughly as follows: First, they divide the interval $[-n, n]$, which contains the first weight of every solution $x \in \mathcal{S}$, into a certain number T of subintervals of length $2n/T$. For $i \in [T]$, let t^T_i denote the i-th threshold, that is, $t^T_i = -n + 2(i-1)n/T$. If the interval length $2n/T$ is so small that none of the intervals contains the first weights of two different solutions from $\mathcal{S}$, then the number of Pareto-optimal solutions equals the number of intervals $[t^T_i, t^T_{i+1}]$ that contain a Pareto-optimal solution. Hence, disregarding some minor technicalities and using the bound on the loser gap stated above, the expected number of Pareto-optimal solutions can be bounded from above by

$$\lim_{T\to\infty} \sum_{i=1}^{T} \mathbf{Pr}\left[\exists x \in \mathcal{P} \colon w^1(x) \in [t^T_i, t^T_{i+1}]\right] =$$

$$\lim_{T\to\infty} \sum_{i=1}^{T} \mathbf{Pr}\left[\Lambda(t^T_i) \leq \frac{2n}{T}\right] \leq \lim_{T\to\infty} \sum_{i=1}^{T} \frac{2n^2\phi}{T} = O(n^2\phi).$$

3.2. Generalized Loser Gap

In this section, we generalize the notion of loser gap. The setup is exactly the same as in the previous section, but instead of just looking at the loser $\hat{x}$ that has the smallest distance to the threshold t, we define for every $u \in \mathbb{N}$, the *u-th loser gap* $\Lambda(t,u)$ as the distance of the first u solutions in $\mathcal{L}$ from the threshold t, i.e., $\Lambda(t,u) = \min\{s \in \mathbb{R} \mid$ there exist u different $x \in \mathcal{L}$ with $w^1(x) \leq s\} - t$.

Naturally, for increasing values of u, the probability that the u-th loser gap is small should decrease, but a priori it is not clear how much we gain by considering more than one loser. The following lemma answers this question.

Lemma 3.1. *For every $\varepsilon \geq 0$, $z \in \mathbb{N}$, and $t \in \mathbb{R}$,*

$$\mathbf{Pr}\left[\Lambda(t, 2^{z-1}) \leq \varepsilon\right] \leq 2^{z^2+z} n^z \phi^z \varepsilon^{z-1}.$$

Proof: Let $u = 2^{z-1}$ and let $x^1, \ldots, x^u$ denote the losers from $\mathcal{L}$ with the u smallest first weights. Then we can write the u-th loser gap $\Lambda(t,u)$ as $\max_{i\in[u]} w^1(x^i) - t$. In Lemma 3.2 we prove that the vectors $x^\star, x^1, \ldots, x^u$ must have rank at least z and that we can find a set $I \subseteq [n]$ of z indices such that the subvectors $x^\star_I, x^1_I, \ldots, x^u_I$ have rank z. We assume without loss of generality that the vectors $x^\star_I, x^1_I, \ldots, x^{z-1}_I$ have rank z and that $x^i = \operatorname{argmin}\{w^1(x) \mid x \in \mathcal{L}, x_I = x^i_I\}$. Then $\Lambda(t,u) \geq \max_{i\in[z-1]} w^1(x^i) - t$.

For every set $I \subseteq [n]$ of z indices and every set $Y = \{y^\star, y^1, \dots, y^{z-1}\} \subseteq \{0,1\}^z$ of linearly independent vectors, we define a random indicator variable $\Lambda_{I,Y} \in \{0,1\}$ below. These random variables have the property that $\Lambda(t,u) \leq \varepsilon$ can only occur if $\Lambda_{I,Y} = 1$ for at least one pair of I and Y. We can thus analyze the random variables $\Lambda_{I,Y}$ and conclude the desired result for $\Lambda(t,u)$ by a union bound.

For given I and Y, we first define $x^\star_{I,Y}$ to be the winner among the solutions that coincide with $y^\star$ in the indices from I, that is, $x^\star_{I,Y} = \operatorname{argmin}\{w^2(x) \mid x \in \mathcal{S}, w^1(x) \leq t, x_I = y^\star\}$. We define the corresponding set of losers as $\mathcal{L}(I,Y) = \{x \in \mathcal{S} \mid w^2(x) < w^2(x^\star_{I,Y})\}$. If no winner $x^\star_{I,Y}$ exists, then $\mathcal{L}(I,Y)$ is defined to be $\mathcal{S}$. The first important observation is that $x^\star_{I,Y}$ coincides with the true winner $x^\star$ if $y^\star = x^\star_I$. Hence, in this case, also $\mathcal{L}(I,Y)$ coincides with the loser set $\mathcal{L}$.

Now we define, for every $i \in [z-1]$, a loser $x^i_{I,Y} = \operatorname{argmin}\{w^1(x) \mid x \in \mathcal{L}(I,Y), x_I = y^i\}$ and the corresponding indicator variable as

$$\Lambda_{I,Y} = \begin{cases} 1 & \text{if } \forall i \in [z-1]\colon w^1(x^i_{I,Y}) \in (t, t+\varepsilon], \\ 0 & \text{otherwise.} \end{cases}$$

Also for the losers the crucial observation is that there is a choice for I and Y such that $x^i_{I,Y}$ coincides with x^i for every $i \in [z-1]$. We just need to choose I so that the vectors $x^\star_I, x^1_I, \dots, x^{z-1}_I$ have rank z, $y^\star = x^\star_I$, and $y^i = x^i_I$ for every $i \in [z-1]$. As mentioned above, the existence of such a choice is guaranteed by Lemma 3.2. The loser gap $\Lambda(t,u)$ is at most ε if and only if, for this choice of I and Y, the random variable $\Lambda_{I,Y}$ takes the value 1.

Let us fix a set I and a set of vectors Y. To analyze the random variable $\Lambda_{I,Y}$, we let an adversary fix all random variables w^1_i with $i \notin I$, that is, we use only the randomness of the variables w^1_i with $i \in I$. Once also the linear combination $w^1_I \cdot y^\star$ of these variables is fixed, the winner $x^\star_{I,Y}$ and hence also the set $\mathcal{L}(I,Y)$ are determined. This implies that the identities of all solutions $x^i_{I,Y}$ are fixed as the (yet unrevealed) weights in w^1_I affect the first weight of all solutions x with $x_I = y^i$ in the same way. Hence, after the weights w^1_i with $i \notin I$ and the linear combination $w^1_I \cdot y^\star$ are fixed, there is a fixed interval $(t^{(i)}, t^{(i)} + \varepsilon]$ of length ε that the linear combination $w^1_I \cdot y^i$ has to fall into in order for $w^1(x^i_{I,Y})$ to be in $(t, t+\varepsilon]$.

Hence, letting an adversary fix all weights w^1_i with $i \notin I$ leaves us with a special case of the scenario analyzed in Lemma 3.3, which yields the following bound:

$$\begin{aligned}\mathbf{Pr}[\Lambda_{I,Y} = 1] &\leq \mathbf{Pr}\left[\forall i \in [z-1]\colon w^1(x^i_{I,Y}) \in (t, t+\varepsilon]\right] \\ &= \mathbf{Pr}\left[\forall i \in [z-1]\colon w^1_I \cdot y^i \in (t^{(i)}, t^{(i)} + \varepsilon]\right] \leq 2z\phi^z\varepsilon^{z-1},\end{aligned}$$

where $t^{(i)}$ is fixed once $w^1_I \cdot y^\star$ is fixed. Now we apply a union bound over all possible choices of I and Y, yielding $\mathbf{Pr}[\Lambda(t,u) \leq \varepsilon] \leq \mathbf{Pr}[\exists I, Y\colon \Lambda_{I,Y} = 1] \leq 2^{z^2} n^z \cdot 2z\phi^z\varepsilon^{z-1} \leq 2^{z^2+z} n^z \phi^z \varepsilon^{z-1}$. ∎

Due to space limitations, the proofs of the following two lemmas are deferred to the full version of this paper.

Lemma 3.2. *For $u \in \mathbb{N}$, let $x^1, \dots, x^u$ be arbitrary distinct vectors from $\{0,1\}^n$. Then the rank r of these vectors when considered as elements of the usual vector space over $\mathbb{R}^n$ is at least $\lceil \log u \rceil$. Furthermore, there exists a set $I \subseteq [n]$ of r indices such that the subvectors $x^1_I, \dots, x^u_I$ have rank r.*

Lemma 3.3. *Let $X_1, \dots, X_r$ be independent random variables, and let $f_i\colon [-1,1] \to [0,\phi]$ denote the density of X_i. Furthermore let $a_1, \dots, a_r \in \{0,1\}^r$ be linearly independent vectors. Let $k \leq r$ and assume that for every $i \in \{r-k+1, \dots, r\}$ there is an arbitrary function $g_i\colon \mathbb{R}^{i-1} \to \mathbb{R}$ given. Denote by $\mathcal{F}(\varepsilon)$ the event that for every $i \in \{r-k+1, \dots, r\}$, the random variable $a_i \cdot X$ takes a value in the interval $[g_i(a_1 \cdot X, \dots, a_{i-1} \cdot X), g_i(a_1 \cdot X, \dots, a_{i-1} \cdot X) + \varepsilon]$, where $X = (X_1, \dots, X_r)$. Then, for every $\varepsilon \geq 0$, $\mathbf{Pr}[\mathcal{F}(\varepsilon)] \leq (2r)^{r-k}\phi^r\varepsilon^k$.*

4. Bicriteria Optimization

In this section, we consider problems with two objective functions. Let us stress that it suffices if one of the objective functions is perturbed. According to our model, $w^2\colon \mathcal{S} \to \mathbb{R}$ can be an arbitrary function that assigns a unique adversarial value to each solution. Only w^1 needs to be linear with perturbed coefficients. The proof of the following theorem is based on the result on the generalized loser gap presented in Lemma 3.1. We know that the first weight of every solution lies in the interval $[-n, n]$. As in the previous analysis [6], we divide this interval into a certain number of subintervals. We then use the generalized loser gap to argue that each of these subintervals can contain only a small number of Pareto-optimal solutions.

Theorem 4.1. *For every $c \in \left[1, \sqrt{\log(n^2\phi)}\right]$ and every sufficiently large value of $n^2\phi$,*

$$\mathbf{E}[q_2^c] \leq \left(n^2\phi\right)^{c\left(1+\frac{5}{\sqrt{\log(n^2\phi)}}\right)} = (n^2\phi)^{c(1+o(1))}.$$

In [7], a lower bound of $\Omega(n^2)$ on the expected number of Pareto-optimal solutions is proven. This immediately implies a lower bound of $\Omega(n^{2c})$ on the expected value of q_2^c for every $c \geq 1$. Hence, our upper bound in Theorem 4.1 almost matches this lower bounds in terms of n, which gives rise to the following concentration result.

Corollary 4.2. *Let $L := \sqrt{\log(n^2\phi)}$ For every $c \in [1, L]$, every sufficiently large value of $n^2\phi$, and $\beta \geq 1$, $\mathbf{Pr}\left[q_2^c \geq \beta \cdot \left(n^2\phi\right)^{c(1+5/L)}\right] \leq \beta^{-L/c}$.*

Proof: We apply Markov's inequality to the L-th moment of q_2: $\mathbf{Pr}\left[q_2^L \geq \beta^{L/c} \cdot \left(n^2\phi\right)^{L(1+5/L)}\right] \leq \beta^{-L/c}$. ∎

Proof of Theorem 4.1: As we assume that the weights w_i^1 take only values in $[-1,1]$, the first weight of every solution lies in the interval $[-n,n]$. We partition this interval into T subintervals of length $2n/T$ for some $T \in \mathbb{N}$ to be chosen later. For $i \in [T]$, let t_i denote the i-th threshold, i.e., $t_i = -n + 2(i-1)n/T$. Let $x_i^\star$ denote the solution from $\mathcal{S}$ that minimizes the second weight $w^2(x)$ under the constraint $w^1(x) \leq t_i$. Let $\mathcal{L}_i$ denote the set of losers as defined in Section 3.2, that is, $\mathcal{L}_i$ is the set of those solutions $x \in \mathcal{S}$ with a smaller second weight than $x_i^\star$. If there is no solution with $w^1(x) \leq t_i$, then we define $\mathcal{L}_i$ to be $\mathcal{S}$.

Observe that all solutions in $\mathcal{L}_i$ must have a first weight larger than t_i. In particular, all Pareto-optimal solutions with a first weight of more than t_i must belong to the set $\mathcal{L}_i$ as they were otherwise dominated by $x_i^\star$. Hence, in order to obtain an upper bound on the number of Pareto-optimal solutions, we can count for every $i \in [T]$, how many losers from $\mathcal{L}_i$ have a first weight between t_i and t_{i+1}, and add up these counts. This implies that the number q_2 of Pareto-optimal solutions can only exceed a given number $t^{1/c}$ if for one $i \in [T]$, the $(t^{1/c}/T)$-th loser gap $\Lambda(t_i, t^{1/c}/T)$ is at most $2n/T$. For $t, z \in \mathbb{N}$ and $T = t^{\frac{1}{c}}/2^{z-1}$, we obtain, by Lemma 3.3,

$$\begin{aligned}\mathbf{Pr}\left[q_2 \geq t^{\frac{1}{c}}\right] &\leq \sum_{i=1}^{T} \mathbf{Pr}\left[\Lambda\left(t_i, \frac{t^{\frac{1}{c}}}{T}\right) \leq \frac{2n}{T}\right] \\ &\leq \sum_{i=1}^{T} \mathbf{Pr}\left[\Lambda\left(t_i, 2^{z-1}\right) \leq \frac{2n}{T}\right] \leq \frac{2^{2z^2} n^{2z-1} \phi^z}{t^{\frac{z-2}{c}}}.\end{aligned}$$

For $z \geq 18$ and $c \in [1, z-3]$, we obtain

$$\begin{aligned}\mathbf{E}[q_2^c] &= \sum_{t=1}^{\infty} \mathbf{Pr}[q_2^c \geq t] \leq \int_{t=0}^{\infty} \mathbf{Pr}[q_2^c \geq t]\ dt \\ &\leq \int_{t=0}^{\infty} \min\left\{1, \frac{2^{2z^2} n^{2z-1} \phi^z}{t^{\frac{z-2}{c}}}\right\} dt \\ &= \frac{z-2}{z-2-c} \cdot \left(2^{2z^2} n^{2z-1} \phi^z\right)^{\frac{c}{z-2}} \\ &\leq \left(2^{3z-9} \cdot n^{2+\frac{3}{z-2}} \cdot \phi^{1+\frac{2}{z-2}}\right)^c,\end{aligned}$$

where we used that $(z-2) \cdot 2^{4+\frac{8}{z-2}} \leq 2^{z-9}$ for $z \geq 18$, and that for every $a > 1$ and $b > 0$, $\int_0^\infty \min\left\{1, \frac{b}{t^a}\right\} dt = \frac{a \cdot b^{1/a}}{a-1}$. We set $z = \sqrt{\log(n^2\phi)} + 3$. Then, we have $c \leq z - 3$, and $z \geq 18$ if $n^2\phi$ is sufficiently large. For this choice of z, we obtain

$$\begin{aligned}\mathbf{E}[q_2^c] &\leq \left(2^{3\sqrt{\log(n^2\phi)}}\right)^c \cdot \left(n^2\phi\right)^{c\left(1+\frac{2}{\sqrt{\log(n^2\phi)}}\right)} \\ &\leq \left(n^2\phi\right)^{c\left(\frac{3}{\sqrt{\log(n^2\phi)}}\right)} \cdot \left(n^2\phi\right)^{c\left(1+\frac{2}{\sqrt{\log(n^2\phi)}}\right)}.\end{aligned}$$

■

5. Multiobjective Optimization

In this section, we analyze the number of Pareto-optimal solutions in multiobjective optimization problems with one adversarial objective function and $d-1$ perturbed linear objective functions. We obtain the following result.

Theorem 5.1. *There are functions $f\colon \mathbb{N} \to \mathbb{N}$ and $g\colon \mathbb{N} \to \mathbb{N}$ such that for every number d of objectives and every $c \in [1, \sqrt{\log(n^2\phi)}/f(d)]$,*

$$\mathbf{E}[q_d^c] \leq \left(n^2\phi\right)^{c\left(f(d)+\frac{g(d)}{\sqrt{\log(n^2\phi)}}\right)} = \left(n^2\phi\right)^{c(f(d)+o(1))}.$$

Unfortunately, the functions f and g grow very fast with d. The bound for f following from our proof is $f(d) = 2^{d-3}d!$, the one for g grows even faster. The reason for this growth is that we use an inductive argument bounding the moments of q_d based on a bound on the moments of q_{d-1}. In each step of this induction, we incur a constant factor in the exponent. The following lemma is the main ingredient in the proof of Theorem 5.1.

Lemma 5.2. *For $d \geq 3$, let $\alpha \in (0,1]$, $A \geq 1$, and $B \geq 1$ be constants, and let $L := \sqrt{\log(n^2\phi)}$. If for every $c \in [1, \alpha L]$ and every sufficiently large $n^2\phi$, $\mathbf{E}\left[q_{d-1}^c\right] \leq \left(n^2\phi\right)^{c\left(A+\frac{B}{L}\right)}$, then for every $c \in [1, \alpha L/(2d)]$ and every sufficiently large $n^2\phi$,*

$$\mathbf{E}[q_d^c] \leq \left(\left(n^2\phi\right)^c\right)^{2dA+\frac{16d^2A+4dB+13}{\alpha L}}.$$

Proof: The setup of the proof is similar to the bicriteria case. As in Theorem 4.1, we partition the interval $[-n,n]$ uniformly into a certain number T of subintervals and consider each of these subintervals separately. For $i \in [T]$, let again t_i denote the i-th threshold, that is, $t_i = -n + 2(i-1)n/T$. In the bicriteria case, the i-th winner was defined to be the solution that minimizes the second weight among all solutions with a first weight of at most t_i. The situation in the multiobjective case is more complicated as there are $d-1$ objectives besides the first one and it is not immediately clear how to generalize the notion of winner appropriately.

For each $i \in [T]$, we consider the set $\mathcal{S}_i$ that consists of the solutions $x \in \mathcal{S}$ with $w^1(x) \leq t_i$. We define $\mathcal{P}_i$ to be the set of solutions from $\mathcal{S}_i$ that are Pareto-optimal among $\mathcal{S}_i$ with respect to the weights $w^2, \ldots, w^d$. The solutions in $\mathcal{P}_i$ take over the role of the winners in the bicriteria case. We associate with every $(d-1)$-tuple $x^\star = (x^{\star,2}, \ldots, x^{\star,d}) \in (\mathcal{P}_i \cup \{\perp\})^{d-1}$ a region $R(x^\star) \subseteq \mathbb{R}^{d-1}$ defined as $R(x^\star) = \{(a_2, \ldots, a_d) \in \mathbb{R}^{d-1} \mid \forall j \in \{2, \ldots, d\}\colon a_j < w^j(x^{\star,j})\}$, where we assume $w^j(\perp) = \infty$ for every $j \in [d]$. If the weight vector $(w^2(x), \ldots, w^d(x))$ of a solution x falls into the region $R(x^\star)$, then x cannot be dominated by any of the solutions $x^{\star,j}$. It can, however, be the case that x is dominated by another solution from $\mathcal{P}_i$. We say that a tuple $x^\star$ is *interesting* if there is no solution $x \in \mathcal{P}_i$ such that the point $(w^2(x), \ldots, w^d(x))$ dominates

$(w^2(x^{\star,2}),\dots,w^d(x^{\star,d}))$. For each interesting tuple $x^\star$, we define a loser set $\mathcal{L}(x^\star)$ as $\mathcal{L}(x^\star) = \{x \in \mathcal{S} \mid \forall j \in \{2,\dots,d\}\colon w^j(x) < w^j(x^{\star,j})\}$.

The following two claims are crucial: First, every Pareto-optimal solution $x \in \mathcal{P}$ with $w^1(x) > t_i$ is contained in the loser set $\mathcal{L}(x^\star)$ for at least one interesting tuple $x^\star \in (\mathcal{P}_i \cup \{\perp\})^{d-1}$. Second, for all interesting tuples $x^\star \in (\mathcal{P}_i \cup \{\perp\})^{d-1}$, all solutions in the loser set $\mathcal{L}(x^\star)$ have a first weight larger than t_i. These claims are proven in the full version of this paper.

For $i \in [T]$, let k_i denote the number of interesting tuples $x^\star \in (\mathcal{P}_i \cup \{\perp\})^{d-1}$. The following observation, which is based on the two claims above, is the essential step in our proof: The union $\mathcal{L}_i$ of the loser sets $\mathcal{L}(x^\star)$ over all interesting tuples $x^\star \in (\mathcal{P}_i \cup \{\perp\})^{d-1}$ contains all Pareto-optimal solutions from $\mathcal{P}$ with a first weight of more than t_i and it contains no solution with a first weight of at most t_i. As for the bicriteria case, we would like to show that there cannot be many solutions in $\mathcal{L}_i$ with a first weight between t_i and t_{i+1}. For this, we define, for each $u \in \mathbb{N}$ and each interesting tuple $x^\star \in (\mathcal{P}_i \cup \{\perp\})^{d-1}$, the u-th loser gap $\Lambda_i(x^\star, u)$ as the distance of the first u losers in $\mathcal{L}(x^\star)$ to the threshold t_i, that is, $\Lambda_i(x^\star,u) = \min\{s \in \mathbb{R} \mid$ there exist u different $x \in \mathcal{L}(x^\star)$ with $w^1(x) \le s\} - t_i$. For $u \in \mathbb{R}_{>0}$, we define $\Lambda_i(x^\star,u) = \Lambda_i(x^\star, \lfloor u \rfloor)$.

Now we can upper bound the number of Pareto-optimal solutions as follows: For each interesting tuple $x^\star \in (\mathcal{P}_i \cup \{\perp\})^{d-1}$, we count the number of losers in $\mathcal{L}(x^\star)$ that have a first weight between t_i and $t_{i+1} = t_i + 2n/T$. As, for every $i \in [T]$, the union of the loser sets $\mathcal{L}(x^\star)$ over all interesting tuples $x^\star \in (\mathcal{P}_i \cup \{\perp\})^{d-1}$ contains all Pareto-optimal solutions from $\mathcal{P}$ with a first weight of more than t_i, the sum of these counts is an upper bound on the number of Pareto-optimal solutions in $\mathcal{P}$. Hence, the number q_d of Pareto-optimal solutions in $\mathcal{P}$ can only exceed a given number $t^{1/c}$ if for one $i \in [T]$, the number of Pareto-optimal solutions with a first weight between t_i and t_{i+1} exceeds $t^{1/c}/T$. This in turn can only happen if, for one $i \in [T]$ and one interesting tuple $x^\star \in (\mathcal{P}_i \cup \{\perp\})^{d-1}$, the $(t^{1/c}/(k_i T))$-th loser gap is at most $2n/T$. Thus $\mathbf{Pr}\left[q_d \ge t^{1/c}\right]$ is bounded from above by

$$\sum_{i=1}^{T} \mathbf{Pr}\left[\exists x^\star \in (\mathcal{P}_i \cup \{\perp\})^{d-1}\colon \Lambda_i\left(x^\star, \frac{t^{1/c}}{k_i T}\right) \le \frac{2n}{T}\right].$$

For $S \in \mathbb{N}$, we can upper bound this term by

$$\sum_{i=1}^{T} \Big(\mathbf{Pr}[k_i \ge S] + \mathbf{Pr}\Big[\exists x^\star \in (\mathcal{P}_i \cup \{\perp\})^{d-1}\colon \Lambda_i\left(x^\star, \frac{t^{1/c}}{ST}\right) \le \frac{2n}{T}\Big]\Big). \quad (1)$$

We cannot directly use Lemma 3.1 to bound the probability that the loser gap Λ_i is small as we have to take into account that each loser set $\mathcal{L}(x^\star)$ is now defined by $d-1$ instead of only one solution. While this requires only minor changes, the main challenge we have to overcome is that the set $\mathcal{P}_i$ is not fixed even when all weight functions except for the first one are chosen. The following lemma shows that these dependencies do not cause too much harm.

Lemma 5.3. *For every $z \in \mathbb{N}$, $S \in \mathbb{N}$, and $\varepsilon \ge 0$, $\mathbf{Pr}\left[\exists x^\star \in (\mathcal{P}_i \cup \{\perp\})^{d-1}\colon \Lambda_i\left(x^\star, 2^z + 2 - d\right) \le \varepsilon\right]$ is bounded from above by*

$$n^z 2^{z^2}\left(\mathbf{Pr}\left[(q_{d-1}+1)^{d-1} \ge S\right] + S \cdot (2z)^{d-1}\phi^z \varepsilon^{z-d+1}\right).$$

Proof: The proof of Lemma 3.1 was only based on the randomness of the first weight function w^1. As the set $\mathcal{P}_i$ is defined to be a set of Pareto-optimal solutions with respect to the other $d-1$ weights, one might think that knowing the set of interesting tuples $x^\star$ does not bias the first weight. Remember, however, that $\mathcal{P}_i$ is defined only among the solutions with a first weight of at most t_i, and hence, the first weight has an impact on the set $\mathcal{P}_i$. Conversely, this implies that the set of interesting tuples and the coefficients in the first weight function are not independent, which we have to take into account when analyzing the probability that there exists an interesting tuple with a small loser gap.

Assume that there is an interesting tuple $x^\star = (x^{\star,2},\dots,x^{\star,d}) \in (\mathcal{P}_i \cup \{\perp\})^{d-1}$ for which the loser gap $\Lambda_i\left(x^\star, 2^{z-1}+2-d\right)$ is at most ε. For $u = 2^{z-1}+2-d$, let $x^1,\dots,x^u$ denote the u losers from $\mathcal{L}(x^\star)$ with the smallest first weights. Together with the winners $x^{\star,2},\dots,x^{\star,d}$, these solutions have rank at least z according to Lemma 3.2. Let $r \le d-1$ denote the rank of the winners, and assume without loss of generality that the vectors $x^{\star,2},\dots,x^{\star,d},x^1,\dots,x^{z-d+1}$ have rank $z' = r+z-d+1$. Then $\Lambda_i\left(x^\star, 2^z+2-d\right) \ge \max_{j\in[z-d+1]} w^1(x^j) - t_i$ and we can find a subset $I \subseteq [n]$ of z' indices and a set $Y = \{y^{\star,2},\dots,y^{\star,d},y^1,\dots,y^{z-d+1}\} \subseteq \{0,1\}^{z'}$ with the following properties: for every $j \in \{2,\dots,d\}$, $x^{\star,j}_I = y^{\star,j}$, and for every $j \in [z-d+1]$, $x^j_I = y^j$ and the vector y^j is linearly independent from the vectors $y^{\star,2},\dots,y^{\star,d},y^1,\dots,y^{j-1}$.

For each such choice of I and Y, we define define $\mathcal{S}_{I,Y}$ to be the set of all solutions $x \in \mathcal{S}$ with $w^1(x) \le t_i$ and $x_I = y^{\star,j}$ for one $j \in \{2,\dots,d\}$. We denote by $\mathcal{P}_{I,Y}$ the set of solutions from $\mathcal{S}_{I,Y}$ that are Pareto-optimal among $\mathcal{S}_{I,Y}$ with respect to the weight functions $w^2,\dots,w^d$. Now for every interesting tuple $x^\star \in (\mathcal{P}_{I,Y} \cup \{\perp\})^{d-1}$, where interesting is defined analogously as for $\mathcal{P}_i$, we define the random indicator variable $\Lambda_{I,Y}(x^\star) \in \{0,1\}$ as follows: For $j \in [z-d+1]$, let $x^j_{I,Y}$ denote the loser from $\mathcal{L}(x^\star)$ with the smallest first weight among all solutions $x \in \mathcal{L}(x^\star)$ with $x_I = y^j$. The random variable $\Lambda_{I,Y}(x^\star)$ is now defined as

$$\begin{cases} 1 & \text{if } \forall j \in [z-d+1]\colon w^1(x^j_{I,Y}) \in (t_i, t_i+\varepsilon], \\ 0 & \text{otherwise.} \end{cases}$$

For fixed I and Y, let $k_{I,Y}$ denote the number of interesting tuples in $(\mathcal{P}_{I,Y} \cup \{\perp\})^{d-1}$. For $S \in \mathbb{N}$, we can

bound the probability of having an interesting tuple $x^\star$ with $\Lambda_{I,Y}(x^\star) = 1$ by the sum of $\mathbf{Pr}[k_{I,Y} \geq S]$ and

$$\mathbf{Pr}\left[\left(\exists x^\star \in (\mathcal{P}_{I,Y} \cup \{\bot\})^{d-1} : \Lambda_{I,Y}(x^\star) = 1\right) \wedge (k_{I,Y} \leq S)\right].$$

In order to analyze the second probability, we use the same arguments as in Lemma 3.3. For this, we first let an adversary fix all weights w_j^1 with $j \notin I$ and all weights w_j^i for $i > 1$ and $j \in [n]$. arbitrarily. If additionally all linear combinations $w_I^1 \cdot y^{\star,j}$ for $j \in [d-1]$ are fixed, then the set of interesting tuples and $k_{I,Y}$ are fixed as well. Hence, as in the proof of Lemma 3.3, we can first integrate over all values the linear combinations $w_I^1 \cdot y^{\star,j}$ can take and for which $k_{I,Y} \leq S$. Once the values for these linear combinations are chosen, we have at most S interesting tuples $x^\star$ in $(\mathcal{P}_{I,Y} \cup \{\bot\})^{d-1}$ and each of them defines a random variable $\Lambda_{I,Y}(x^\star)$. As in Lemma 3.3, this random variable can only be 1 if every linear combination $w_I^1 \cdot y^j$, for $j \in [z-d+1]$, takes a value in a fixed interval of length ε. Hence, when taking into account that we have at most S interesting tuples, a union bound combined with the same reasoning as in Lemma 3.3 implies

$$\begin{aligned}&\mathbf{Pr}\left[\left(\exists x^\star \in (\mathcal{P}_{I,Y} \cup \{\bot\})^{d-1} : \Lambda_{I,Y}(x^\star) = 1\right) \wedge (k_{I,Y} \leq S)\right] \\ &\leq S \cdot (2z)^{d-1} \phi^{z'} \varepsilon^{z'-d+1} \leq S \cdot (2z)^{d-1} \phi^z \varepsilon^{z-d+1}.\end{aligned}$$

Let us now return to the assumption that there is an interesting tuple $x^\star = (x^{\star,2}, \ldots, x^{\star,d}) \in (\mathcal{P}_i \cup \{\bot\})^{d-1}$ for which the loser gap $\Lambda_i(x^\star, u)$ is at most ε. Lemma 3.2 guarantees that we can choose I and Y such that $x^\star \in (\mathcal{P}_{I,Y} \cup \{\bot\})^{d-1}$ and such that, for every $j \in [z-d+1]$, the loser x^j coincides with $x_{I,Y}^j$. Hence, there can only exist an interesting tuple $x^\star$ whose loser gap $\Lambda_i(x^\star, u)$ is smaller than ε if for one choice of I, Y, and $x^\star \in (\mathcal{P}_{I,Y} \cup \{\bot\})^{d-1}$ the random variable $\Lambda_{I,Y}(x^\star)$ takes the value 1. Thus, a union bound over all choices for I and Y implies

$$\begin{aligned}&\mathbf{Pr}\left[\exists x^\star \in (\mathcal{P}_i \cup \{\bot\})^{d-1} \colon \Lambda_i(x^\star, 2^z + 2 - d) \leq \varepsilon\right] \leq \\ &n^z 2^{z^2} \left(\mathbf{Pr}\left[(q_{d-1}+1)^{d-1} \geq S\right] + S \cdot (2z)^{d-1} \phi^z \varepsilon^{z-d+1}\right).\end{aligned}$$

■

The number k_i of interesting tuples $x^\star \in (\mathcal{P}_i \cup \{\bot\})^{d-1}$ is bounded by $(|\mathcal{P}_i|+1)^{d-1} \leq (2|\mathcal{P}_i|)^{d-1}$. For $S = \frac{t^{1/c}}{T(2^z+2-d)}$, combining (1) with Lemma 5.3 yields the following upper bound on $\mathbf{Pr}\left[q_d \geq t^{1/c}\right]$:

$$\begin{aligned}&\sum_{i=1}^{T} \left(\mathbf{Pr}\left[q_{d-1}^{d-1} \geq \frac{S}{2^{d-1}}\right] + n^z 2^{z^2} \left(\mathbf{Pr}\left[q_{d-1}^{d-1} \geq \frac{S}{2^{d-1}}\right]\right.\right. \\ &\left.\left. + \; S \cdot (2z)^{d-1} \phi^z \left(\frac{2n}{T}\right)^{z-d+1}\right)\right) \\ \leq \quad &T \cdot n^z 2^{z^2+1} \cdot \mathbf{Pr}\left[q_{d-1}^{d-1} \geq \frac{t^{1/c}}{T 2^{d-1}(2^z+2-d)}\right] \\ &+ \frac{t^{1/c} \cdot 2^{z^2+zd} n^{2z-d+1} \phi^z}{T^{z-d+1}(2^z+2-d)}. \qquad (2)\end{aligned}$$

In order to estimate the probability of q_{d-1}^{d-1} being too large, we use a simple concentration result that follows from Markov's inequality when applied to the (αL)-th moment of q_{d-1}:

$$\mathbf{Pr}\left[q_{d-1}^{d-1} \geq \beta (n^2\phi)^{(d-1)\left(A+\frac{B}{L}\right)}\right] \leq \beta^{-\frac{\alpha L}{d-1}}. \qquad (3)$$

We set $z = \alpha L$ and, for some $k \in (0,1)$ to be chosen later, $T = \frac{t^{\frac{1-k}{c}}}{2^{d-1}(2^z+2-d)(n^2\phi)^{(d-1)\left(A+\frac{B}{L}\right)}}$, which yields combined with (2) and (3) the following bound:

$$\begin{aligned}\mathbf{Pr}\left[q_d \geq t^{1/c}\right] &\leq \frac{T \cdot n^z 2^{z^2+1}}{t^{\frac{kz}{c(d-1)}}} + \frac{t^{1/c} \cdot 2^{z^2+zd} n^{2z-d+1} \phi^z}{T^{z-d+1}(2^z+2-d)} \\ &\leq \frac{2n^{z+2}\phi}{t^{\frac{kz+(k-1)(d-1)}{c(d-1)}}} + \frac{(n^2\phi)^{zdA+dB+3}}{t^{\frac{(1-k)(z-d+1)-1}{c}}}.\end{aligned}$$

For $k = (\alpha L + 2d - 2)/(2\alpha L + 2d - 2)$, we obtain, for large enough $n^2\phi$,

$$\mathbf{Pr}\left[q_d \geq t^{1/c}\right] \leq \frac{(n^2\phi)^{\alpha L}}{t^{\frac{\alpha L}{2c(d-1)}}} + \frac{(n^2\phi)^{\alpha L dA + dB + 3}}{t^{\frac{\alpha L - 2d}{2c}}}.$$

Since we assume $c \leq \alpha L/(2d)$ and $d \geq 3$, both exponents of t are larger than 1 for large enough $n^2\phi$. In the following, we use that for every $a > 1$ and $b > 0$, $\int_0^\infty \min\left\{1, \frac{b}{t^a}\right\} dt = \frac{a \cdot b^{1/a}}{a-1}$. We obtain

$$\begin{aligned}\mathbf{E}[q_d^c] &\leq \int_{t=0}^{\infty} \mathbf{Pr}\left[q_d \geq t^{1/c}\right] dt \\ &\leq \int_{t=0}^{\infty} \min\left\{\frac{(n^2\phi)^{\alpha L}}{t^{\frac{\alpha L}{2c(d-1)}}} + \frac{(n^2\phi)^{\alpha L dA + dB + 3}}{t^{\frac{\alpha L - 2d}{2c}}}, 1\right\} dt \\ &\leq d(n^2\phi)^{2c(d-1)} + 2((n^2\phi)^c)^{2dA + \frac{8d^2A + 2dB + 6)}{\alpha L - 2d}} \\ &\leq ((n^2\phi)^c)^{2dA + \frac{16d^2A + 4dB + 13}{\alpha L}}.\end{aligned}$$

■

Now Theorem 5.1 follows easily by induction on d, using Theorem 4.1 as a start and Lemma 5.2 as the induction step.

6. Extensions and Applications

6.1. Generalized Winner Gap

Another structural property that has proven useful in previous analyses is the *winner gap* Δ. Let us consider an arbitrary single-criterion optimization problem in which a linear objective function $w(x) = w_1 x_1 + \cdots + w_n x_n$ is to be minimized over a feasible region $\mathcal{S} \subseteq \{0,1\}^n$. The winner gap is defined to be the distance between the optimal solution and the second best solution in terms of the objective function w. That is, if $x^{\star 1}$ denotes the optimal solution and $x^{\star 2}$ denotes the second best solution, then $\Delta = w(x^{\star 2}) - w(x^{\star 1})$.

As for the loser gap, we define a generalization of the winner gap that does not only take into account the second best solution, but also the third best solution and so on. We define the *u-th winner gap* $\Delta(u)$ to be the distance

between the optimal solution and the u-th best solution. That is, if $x^{\star u}$ denotes the solution from $\mathcal{S}$ with the u-th smallest objective value $w(x)$, then $\Delta(u) = w(x^{\star u}) - w(x^{\star 1})$. Similar arguments as in the proof of Lemma 3.1 yield the following lemma.

Lemma 6.1. *For every $\varepsilon \geq 0$ and $z \in \mathbb{N}$,* $\mathbf{Pr}\left[\Delta(2^{z-1}+1) \leq \varepsilon\right] \leq 2^{z^2+z} n^z \phi^z \varepsilon^{z-1}$.

6.2. Expected Polynomial Running Time

In this section, we prove the following result: any binary optimization problem that can be solved in randomized pseudo-polynomial time in the worst case can be solved in expected polynomial time in the model of smoothed analysis. This strengthens Beier and Vöcking's characterization, which says that every such problems can be solved with reasonable probability in polynomial time on smoothed instances [8]. For the proof, we combine the adaptive rounding scheme proposed in [8] with our notion of generalized winner and loser gap.

To keep the presentation simple, let us focus on single-criterion problems that have a perturbed linear objective function: We consider optimization problems Π in whose instances a linear objective function $w(x) = w_1 x_1 + \cdots + w_n x_n$ is to be minimized over an arbitrary feasible region $\mathcal{S} \subseteq \{0,1\}^n$. The problem Π could, for example, be the traveling salesperson problem and the coefficients w_i could be the edge lengths. We denote by Π_u the version of Π in which the coefficients w_i are encoded in unary. If $\Pi_u \in \mathrm{ZPP}$, then we say that Π can be solved in randomized pseudo-polynomial time. With other words, if $\Pi_u \in \mathrm{ZPP}$, then there exists a randomized algorithm for Π whose worst-case expected running time is polynomial in the input size N and the largest coefficient $W = \max_i |w_i|$.

Now we consider the smoothed version of Π in which the coefficients w_i are independent random variables distributed according to densities $f_i\colon [-1,1] \to [0,\phi]$. Beier and Vöcking show that if and only if $\Pi_u \in \mathrm{ZPP}$, there is an algorithm A for the smoothed version of Π with the following property: there is a polynomial p such that the probability that the running time of A exceeds $p(N, \phi, 1/\varepsilon)$ is bounded from above by ε for every input of size N, every $\phi \geq 1$, and every $\varepsilon \geq 0$.

One has to be a bit careful, when defining the running time on smoothed instances, as the coefficients w_i are assumed to be continuous random variables, which means that with probability 1 they do not have a finite encoding. To circumvent this problem, Beier and Vöcking introduce an oracle model, in which the bits after the binary point of the coefficients w_i are revealed from left to right by an oracle. Each oracle query needs constant time and yields one more bit of each coefficients w_i. To avoid the somewhat cumbersome oracle model, one can also assume that the coefficients w_i are discretized by rounding them after a polynomial number, say n^2, of bits.

We prove the following characterization theorem.

Theorem 6.2. *A binary optimization problem Π can be solved in expected polynomial time in the model of smoothed analysis if and only if $\Pi_u \in \mathrm{ZPP}$.*

Proof Sketch: As a first ingredient of the proof, we use that if we can find the optimal solution of a binary optimization problem with a linear objective function efficiently, then we can also find the best z solutions, for any constant z, efficiently. The proof of the fact can be found in the full version of this paper.

First, we discuss how an algorithm A with randomized pseudo-polynomial running time can be transformed into an algorithm with expected polynomial running time in the smoothed model. There exists a constant $\ell \geq 1$ such that the expected running time of algorithm A is bounded from above by $(NW)^\ell$. We claim that the following algorithm has expected polynomial running time on smoothed instances:

1) Set $b = 1$ and $u = 2^{\ell+1} + 1$.
2) Round every coefficient w_i after b bits after the binary point. Denote by $\lfloor w_i \rfloor_b$ the rounded coefficients and by $\lfloor w \rfloor_b$ the corresponding vector.
3) Compute the u best solutions $x_b^{\star 1}, \ldots, x_b^{\star u}$ with respect to the rounded objective function $\lfloor w \rfloor_b \cdot x$.
4) If with respect to the rounded coefficients, $x_b^{\star 1}$ has an objective value that is better by at most $n2^{-b}$ than the objective value of $x_b^{\star u}$, then increase b by one and go to step 2.
5) From the u computed solutions $x_b^{\star 1}, \ldots, x_b^{\star u}$ in the previous step output the one with minimal objective value with respect to the original, unrounded objective function $w \cdot x$.

We need to prove two claims: First, the expected running time of this algorithm is polynomial. Second, the algorithm always outputs the optimal solution to the problem with unrounded coefficients. Due to space limitations, the proofs of these claims are deferred to the full version of this paper. ■

The model considered in [8] also considers the case that instead of the objective function, a linear constraint is perturbed. In fact, also our stronger characterization can be extended to that case, when we assume that the (adversarial) objective function is linear. Then we can use the generalized loser gap instead of the generalized winner gap to obtain a similar result as in Theorem 6.2.

6.3. Extension to Integer Optimization Problems

So far, we have considered the case of binary optimization problems, in which the feasible region $\mathcal{S}$ is a subset of $\{0,1\}^n$. This case covers already most of the interesting combinatorial optimization problems. Let us, however, mention that our results can also be extended to the case that the feasible region is a subset of Z^n for a finite set of integers $Z \subseteq \mathbb{Z}$. The important parameter is then the

maximum absolute value of the integers in Z, i.e., $m := \max_{z \in Z} |z|$. Without significant changes to our proofs, we can replace the two occurrences of $\log(n^2\phi)$ in Theorem 5.1 by $\log_{m+1}(n^2\phi)$. Hence, if $\log_{m+1}(n^2\phi) \to \infty$, there is no qualitative difference compared to the binary case. Otherwise, the proof can be adapted to yield a function $h\colon \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ such that for every number d of objectives and every $c \in \mathbb{N}$, the expected value of q_d^c is bounded from above by $(mn\phi)^{h(d,c)}$.

7. Discussion and Open Problems

We introduced a novel technique to analyze multiobjective optimization problems in the model of smoothed analysis. This technique gives not only the first bounds for the smoothed number of Pareto-optimal solutions in multiobjective optimization, but also the first concentration bounds for this quantity. It can also be applied to transform pseudo-polynomial time algorithms into algorithms with expected polynomial running time on smoothed instances.

However, there is still a variety of open questions in the realm of smoothed analysis of multiobjective optimization. The most apparent question is probability whether the exponent in Theorem 5.1 can be improved to a polynomial in d. Another question is whether our method can also be used to bound higher moments. The bound of $\sqrt{\log n^2\phi}$ that we currently have in Lemma 4.1 is due to the occurrence of 2^{z^2} in the bound for the loser gap in Lemma 3.1. If this could be improved to, e.g., $2^{z \log z}$, then we would obtain a bound for even higher moments, leading to stronger concentration bounds. Finally, lower bounds on the smoothed number of Pareto-optimal solutions and an improved analysis of the integer case would be of interest.

References

[1] H. Ackermann, A. Newman, H. Röglin, and B. Vöcking, "Decision-making based on approximate and smoothed Pareto curves." *Theoretical Computer Science*, vol. 378, no. 3, pp. 253–270, 2007.

[2] D. Arthur, B. Manthey, and H. Röglin, "k-means has polynomial smoothed complexity," in *Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2009.

[3] D. Arthur and S. Vassilvitskii, "Worst-case and smoothed analysis of the ICP algorithm, with an application to the k-means method," in *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2006, pp. 153–164.

[4] L. Becchetti, S. Leonardi, A. Marchetti-Spaccamela, G. Schäfer, and T. Vredeveld, "Average case and smoothed competitive analysis of the multilevel feedback algorithm," *Mathematics of Operations Research*, vol. 31, no. 1, pp. 85–108, 2006.

[5] R. Beier, "Probabilistic analysis of discrete optimization problems," Ph.D. dissertation, Universität des Saarlandes, 2004.

[6] R. Beier, H. Röglin, and B. Vöcking, "The smoothed number of Pareto optimal solutions in bicriteria integer optimization," in *Proceedings of the 12th International Conference on Integer Programming and Combinatorial Optimization (IPCO)*, 2007, pp. 53–67.

[7] R. Beier and B. Vöcking, "Random knapsack in expected polynomial time," *Journal of Computer and System Sciences*, vol. 69, no. 3, pp. 306–329, 2004.

[8] ——, "Typical properties of winners and losers in discrete optimization," *SIAM Journal on Computing*, vol. 35, no. 4, pp. 855–881, 2006.

[9] H. W. Corley and I. D. Moon, "Shortest paths in networks with vector weights," *Journal of Optimization Theory and Application*, vol. 46, no. 1, pp. 79–86, 1985.

[10] I. Diakonikolas and M. Yannakakis, "Small approximate pareto sets for bi-objective shortest paths and other problems," in *Proceedings of the 10th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX)*, 2007, pp. 74–88.

[11] M. Ehrgott, "Integer solutions of multicriteria network flow problems," *Investigacao Operacional*, vol. 19, pp. 229–243, 1999.

[12] ——, *Multicriteria Optimization*, ser. Lecture Notes in Economics and Mathematical Systems. Springer, 2000.

[13] M. Englert, H. Röglin, and B. Vöcking, "Worst case and probabilistic analysis of the 2-Opt algorithm for the TSP," in *Proceedings of the 18th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2007, pp. 1295–1304.

[14] P. Hansen, "Bicriterion path problems," in *Multiple Criteria Decision Making: Theory and Applications*, ser. Lecture Notes in Economics and Mathematical Systems, vol. 177, 1980, pp. 109–127.

[15] K. Klamroth and M. M. Wiecek, "Dynamic programming approaches to the multiple criteria knapsack problem," *Naval Research Logistics*, vol. 47, no. 1, pp. 57–76, 2000.

[16] M. Müller-Hannemann and K. Weihe, "Pareto shortest paths is often feasible in practice." in *Proceedings of the 5th International Workshop on Algorithm Engineering (WAE)*, 2001, pp. 185–198.

[17] A. Mustafa and M. Goh, "Finding integer efficient solutions for bicriteria and tricriteria network flow problems using dinas." *Computers & Operations Research*, vol. 25, no. 2, pp. 139–157, 1998.

[18] G. L. Nemhauser and Z. Ullmann, "Discrete dynamic programming and capital allocation," *Management Science*, vol. 15, no. 9, pp. 494–505, 1969.

[19] C. H. Papadimitriou and M. Yannakakis, "On the approximability of trade-offs and optimal access of web sources," in *Proceedings of the 41st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2000, pp. 86–92.

[20] A. J. V. Skriver and K. A. Andersen, "A label correcting approach for solving bicriterion shortest-path problems." *Computers & Operations Research*, vol. 27, no. 6, pp. 507–524, 2000.

[21] D. A. Spielman and S.-H. Teng, "Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time." *Journal of the ACM*, vol. 51, no. 3, pp. 385–463, 2004.

[22] S. Vassilvitskii and M. Yannakakis, "Efficiently computing succinct trade-off curves." *Theoretical Computer Science*, vol. 348, no. 2–3, pp. 334–356, 2005.

APPENDIX

1. Proofs from Section 3

Proof of Lemma 3.2: Assume without loss of generality that the vectors $x^1, \dots, x^r$ are a basis of the space spanned by $x^1, \dots, x^u$. Let M denote the $n \times r$ matrix whose columns are the vectors $x^1, \dots, x^r$. As the rank of this matrix is r, we can find a set $I \subseteq [n]$ of r indices such that the rows corresponding to I are linearly independent and form a basis of the row space of M.

We claim that for every vector $y \in \{0,1\}^r$ there is at most one vector x^i with $x^i_I = y$. As the vectors $x^1_I, \dots, x^r_I$ form a basis of $\mathbb{R}^r$, there exists, for every $y \in \{0,1\}^r$, a unique set of coefficients $\lambda_1, \dots, \lambda_r \in \mathbb{R}$ such that $\lambda_1 x^1_I + \dots + \lambda_r x^r_I = y$. This means that it is not possible to obtain two different vectors as linear combinations of $x^1, \dots, x^r$ that coincide on the indices in I. As the vectors $x^1, \dots, x^r$ are a basis of the space spanned by $x^1, \dots, x^u$, this proves that for every vector $y \in \{0,1\}^r$ there is at most one vector x^i with $x^i_I = y$. Hence, $u \leq 2^r$ and $r \geq \lceil \log u \rceil$. Moreover, we have identified a set $I \subseteq [n]$ with $|I| = r$ such that $x^1_I, \dots, x^r_I$ have rank r. ∎

Proof of Lemma 3.3: As the vectors $a_1, \dots, a_r$ are linearly independent, they form a basis of $\mathbb{R}^r$. We perform a change of variables from the canonical basis to $(a_1, \dots, a_r)$. Let $\Phi \colon \mathbb{R}^r \to \mathbb{R}^r$ denote the corresponding transformation. For the matrix A whose columns are the vectors $a_1, \dots, a_r$, we can write Φ as $\Phi(x) = Ax$.

For $i \in \{r-k+1, \dots, r\}$, we use $I_i(a_1 \cdot X, \dots, a_{i-1} \cdot X)$ to denote the interval

$$[g_i(a_1 \cdot X, \dots, a_{i-1} \cdot X), g_i(a_1 \cdot X, \dots, a_{i-1} \cdot X) + \varepsilon].$$

If we denote by f the joint density of the random variables $X_1, \dots, X_r$ and set $y_{<i} := (y_1, \dots, y_{i-1})$, then we can write the probability of the event $\mathcal{F}(\varepsilon)$ as

$$\begin{aligned}
&\int_{y_1 \in \mathbb{R}} \cdots \int_{y_{r-k} \in \mathbb{R}} \int_{y_{r-k+1} \in I_i(y_{<r-k+1})} \cdots \\
&\quad \int_{y_r \in I_r(y_{<r})} \frac{1}{|\det A|} \cdot f(\Phi^{-1}(y))\, dy_r \cdots dy_1 \\
&\leq \frac{\phi^r}{|\det A|} \int_{y_1 \in [-r,r]} \cdots \int_{y_{r-k} \in [-r,r]} \\
&\quad \int_{y_{r-k+1} \in I_i(y_{<r-k+1})} \cdots \int_{y_r \in I_r(y_{<r})} 1\, dy_r \cdots dy_1,
\end{aligned}$$

where the inequality follows because every random variable $a_i \cdot X$ can only take values in $[-r, r]$. As all the intervals $I_i(y_{<i})$ have length ε, we obtain the desired bound of

$$\frac{\phi^r (2r)^{r-k} \varepsilon^k}{|\det A|} \leq (2r)^{r-k} \phi^r \varepsilon^k$$

because the determinant of A is an integer and not equal to zero. ∎

2. Proofs from Section 5

Proof of the claims in the proof of Theorem 5.1: Let us consider the first claim. For this, let $x \in \mathcal{P}$ be an arbitrary Pareto-optimal solution with $w^1(x) > t_i$. For $j \in \{2, \dots, d\}$, we set $x^{\star,j} = \operatorname{argmin}\{w^j(y) \mid y \in \mathcal{P}_i,\, w^j(x) < w^j(y)\}$ and $x^{\star,j} = \perp$ if there exists no solution $y \in \mathcal{P}_i$ with $w^j(x) < w^j(y)$. We claim that for these choices, $x \in \mathcal{L}(x^\star)$. The tuple $x^\star$ must be interesting because otherwise, if there was a solution $y \in \mathcal{P}_i$ whose weight vector dominates $(w^2(x^{\star,2}), \dots, w^d(x^{\star,d})) \in \mathbb{R}^{d-1}$, then either this solution also dominates x, contradicting the assumption that x is Pareto-optimal. Or, if y does not dominate x, then since $w^1(y) \leq t_i < w^1(x)$, there must exist an index $j \in \{2, \dots, d\}$ with $w^j(x) < w^j(y) < w^j(x^{\star,j})$, contradicting our choice of $x^{\star,j}$. Here we implicitly used the fact that in our probabilistic model, with probability one, there are no two solutions with exactly the same weight. Hence, $x^\star$ is interesting and from its definition it follows immediately that $x \in \mathcal{L}(x^\star)$.

Now let us consider the second claim. Assume for contradiction that there exist an interesting tuple $x^\star$ and a solution $x \in \mathcal{L}(x^\star)$ with $w^1(x) \leq t_i$. As $x^\star$ is interesting, there is no solution $y \in \mathcal{S}$ with $w^1(y) \leq t_i$ for which the point $(w^2(y), \dots, w^d(y))$ dominates the point $(w^2(x^{\star,2}), \dots, w^d(x^{\star,d}))$. Since by the definition of the loser set $(w^2(x), \dots, w^d(x)) < (w^2(x^{\star,2}), \dots, w^d(x^{\star,d}))$, this implies that x cannot be dominated by any solution y with $w^1(y) \leq t_i$. As we assume that $w^1(x) \leq t_i$, this implies that y cannot be dominated by any solution from $\mathcal{S}$ and must hence be Pareto-optimal. This in turn implies that $x \in \mathcal{P}_i$ and hence, $x^\star$ cannot be an interesting tuple as $(w^2(x^{\star,2}), \dots, w^d(x^{\star,d}))$ is dominated by $(w^2(x), \dots, w^d(x))$. ∎

3. Proofs from Section 6.2

Proof of Lemma 6.1: We set $u = 2^{z-1}+1$ and consider the solutions $x^{\star 1}, \dots, x^{\star u}$ with the u smallest objective values. These solutions must have rank at least z according to Lemma 3.2. Without loss of generality, we assume that we can find a subset $I \subseteq [n]$ of z indices such that the vectors $x^{\star 1}_I, \dots, x^{\star z}_I$ have rank z.

For every subset $I \subseteq [n]$ of size z and every set $\{y^1, \dots, y^z\} \subseteq \{0,1\}^z$ of linearly independent vectors, we define a random indicator variable $\Delta_{I,Y}$ as follows: Let $x^{\star i}_{I,Y}$ denote the solution with the smallest objective value among all solutions $x \in \mathcal{S}$ with $x_I = y^i$, and define $\Delta_{I,Y}$ to be

$$\begin{cases} 1 & \text{if } \forall i \in \{2, \dots, z\} \colon w(x^{\star 1}_{I,Y}) - w(x^{\star i}_{I,Y}) \in [0, \varepsilon], \\ 0 & \text{otherwise.} \end{cases}$$

If we choose I such that $x^{\star 1}_I, \dots, x^{\star z}_I$ have rank z and if we choose Y such that $x^{\star i}_I = y^i$ for all $i \in [z]$, then the solutions $x^{\star i}_{I,Y}$ coincide with the solutions $x^{\star i}$. Hence, if the

winner gap is at most ε, then there is a choice for I and Y such that $\Delta_{I,Y} = 1$.

Let us now fix I and Y and analyze $\Delta_{I,Y}$. We allow an adversary to fix all weights w_i for $i \notin I$. If additionally the linear combination $w_I \cdot y^1$ is fixed, then for each $i \in \{2, \dots, z\}$, there is an interval of length ε that the linear combination $w_I \cdot y^i$ has to fall into in order for $\Delta_{I,Y}$ to be 1. Hence, we can apply Lemma 3.3, to obtain an upper bound of $2z\phi^z\varepsilon^{z-1}$ for the probability of $\Delta_{I,Y}$ being 1. A union bound over all choices for I and Y yields the lemma. ∎

Lemma A.1. *If for a problem Π there is a (randomized) algorithm A with (expected) running time $T(N, W)$ for solving instances of input size N with largest number W, then there exists a (randomized) algorithm with (expected) running time $(2n)^{z^2} \cdot T(N, nW) + \mathrm{poly}(N, W)$ for finding the z best solutions, for any constant z.*

Proof: Let $x^1, \dots, x^z$ denote the z best solutions. For each pair $\{x^a, x^b\}$, there must be an index i, with $x_i^a \neq x_i^b$. Let I with $|I| \leq z^2$ denote the set of these indices. Assume that the set I is known. Then we can, for every $y \in \{0,1\}^{|I|}$, use algorithm A to compute the optimal solution $x^\star(y)$ among all solutions $x \in \mathcal{S}$ that satisfy $x_I = y$. This can be accomplished by modifying the coefficients in the objective function as follows: We set $w_i = nW$ for all $i \in I$ with $y_i = 0$ and $w_i = -nW$ for all $i \in I$ with $y_i = 1$. The solution that is optimal with respect to this modified objective function must satisfy $x_I = y$ if there exists a solution $x \in \mathcal{S}$ with $x_I = y$. Hence, the optimal solution of this modified instance is $x^\star(y)$.

This implies that if we have chosen the right index set I, then the best z solutions are among the solutions $x^\star(y)$. We try all possibilities for choosing the set I. This gives us in total at most $(2n)^{z^2}$ different solutions $x^\star(y)$. We just need to output the best z of them. ∎

Additions to the Proof of Theorem 6.2: The correctness follows easily: When the algorithm stops, then all solutions that are not among the solutions $x_b^{\star 1}, \dots, x_b^{\star u}$ have, with respect to the rounded coefficients, a distance of more than $n2^{-b}$ from the rounded value of $x_b^{\star 1}$. Since rounding lowers the value of each solution by at most $n2^{-b}$, none of these solutions can be better than $x_b^{\star 1}$ with respect to the original coefficients. Hence, the optimal solution must be among the solutions $x_b^{\star 1}, \dots, x_b^{\star u}$.

Now let us consider the expected running time: For fixed b, the running time of one iteration is bounded by $(N2^b)^\ell$ as every rounded number can be represented by b bits after the binary point. That is, when we scale all numbers by a factor of 2^b, then we obtain an instance with integer coefficients that are bounded by 2^b. It can only happen that the accuracy of b bits is not enough if the distance between $x_b^{\star 1}$ and all $x_b^{\star i}$ for $i \in [a]$ in terms of the rounded objective function $\lfloor w \rfloor_b \cdot x$ is at most $n2^{-b}$. In this case, in terms of the original objective function $w \cdot x$, the values of the solutions $x_b^{\star i}$ can be at most $n2^{-b+1}$ away from the value $w \cdot x_b^{\star 1}$. Also in terms of the original objective function, the true winner $x^\star$ can be better than $x_b^{\star 1}$ by at most $n2^{-b}$. Hence the distance between $x^\star$ and the solutions $x_b^{\star i}$ in terms of the original coefficients is at most $n2^{-b+2}$. This can only happen if the u-th winner gap is smaller than $n2^{-b+2}$, which happens according to Lemma 6.1 with probability at most $\kappa n^{\ell+2}\phi^{\ell+2}(n2^{-b+2})^{\ell+1}$, for a sufficiently large constant κ depending only on ℓ. Hence, we obtain the following bound on the running time T of algorithm A:

$$\begin{aligned}\mathbf{E}[T] &\leq \sum_{b=1}^{\infty} (N2^b)^\ell \cdot \mathbf{Pr}\left[\Delta(u) \leq n2^{-b+2}\right] \\ &\leq \sum_{b=1}^{\infty} (N2^b)^\ell \cdot \kappa n^{\ell+2}\phi^{\ell+2}(n2^{-b+2})^{\ell+1}) \\ &\leq \mathrm{poly}(N, \phi) \cdot \sum_{b=1}^{\infty} \frac{1}{2^b} = \mathrm{poly}(N, \phi).\end{aligned}$$

The other direction follows from the previous characterization: If there exists an algorithm with expected polynomial running time in the smoothed model, then this algorithm also has polynomial smoothed complexity in the weaker sense used in [8]. Hence, according to Beier and Vöcking's characterization, the problem can be solved in randomized pseudo-polynomial time in the worst case. ∎

4. A Lemma about Random Vectors

In this section, we present a cute lemma about random vectors that follows from our result about the expected number of Pareto-optimal solutions. We find this lemma interesting on its own, but so far we have neither found it in the literature nor have we found a direct way to prove it.

Lemma A.2. *Let $W^1, \dots, W^n \in [-1, 1]^d$ be random vectors and assume that every coordinate of every vector is an independent random variable drawn uniformly at random from $[-1, 1]$. For every constant d, the probability that there exists a subset $I \subseteq [n]$ of these vectors such that $\sum_{i \in I} W^i > \vec{0}$ is at least $1 - \mathrm{poly}(n)/2^n$.*

Proof: Consider a binary optimization problem with $\mathcal{S} = \{0,1\}^n$ and d linear objective functions $w^1, \dots, w^d$. Assume that the coefficients in all objective functions are chosen uniformly at random from $[-1, 1]$. The probability that the solution 0^n is not Pareto-optimal coincides with the probability in the statement of the lemma that we want to bound. To see this, one has to use the fact that there exists only with probability zero a non-empty subset I such that one of the components of the vector $\sum_{i \in I} W^i$ is exactly zero.

We claim that the probability that a solution $x \in \{0,1\}^n$ is Pareto-optimal is exactly the same for every solution x.

The solution x is Pareto-optimal if there does not exist a subset $I \subseteq [n]$ such that for every $i \in [d]$, the sum $\sum_{j \in I, x_j = 0} w_j^i - \sum_{j \in I, x_j = 1} w_j^i$ is at least 0. (If x is dominated by a solution y, then I is the set of indices in which x and y differ.) Since the coefficients are chosen uniformly in $[-1, 1]$, the probability p of not being Pareto-optimal is the same for every x. Hence, the expected number of Pareto-optimal solutions is $2^n \cdot (1 - p)$. Using Theorem 5.1 yields $2^n \cdot (1 - p) = \mathrm{poly}(n)$, proving the lemma. ■