FULL LENGTH PAPER

Series A

The smoothed number of Pareto-optimal solutions in bicriteria integer optimization

René Beier[1] · Heiko Röglin[2] · Clemens Rösner[2] · Berthold Vöcking[3]

© The Author(s) 2022

Abstract

A well-established heuristic approach for solving bicriteria optimization problems is to enumerate the set of Pareto-optimal solutions. The heuristics following this principle are often successful in practice. Their running time, however, depends on the number of enumerated solutions, which is exponential in the worst case. We study bicriteria integer optimization problems in the model of smoothed analysis, in which inputs are subject to a small amount of random noise, and we prove an almost tight polynomial bound on the expected number of Pareto-optimal solutions. Our results give rise to tight polynomial bounds for the expected running time of the Nemhauser-Ullmann algorithm for the knapsack problem and they improve known results on the running times of heuristics for the bounded knapsack problem and the bicriteria shortest path problem.

✉ Heiko Röglin
roeglin@cs.uni-bonn.de

René Beier
Rene.Beier@tomtom.com

Clemens Rösner
roesner@cs.uni-bonn.de

[1] Max-Planck-Institut für Informatik, Saarbrücken, Germany

[2] Department of Computer Science, University of Bonn, Bonn, Germany

[3] Department of Computer Science, RWTH Aachen University, Aachen, Germany

1 Introduction

We study integer optimization problems with two objective functions (say profit and weight) that are to be optimized simultaneously. A common approach for solving such problems is generating the set of *Pareto-optimal* solutions, also known as the *Pareto set*. Pareto-optimal solutions are optimal compromises of the two criteria in the sense that any improvement of one criterion implies an impairment of the other. In other words, a solution x is Pareto-optimal if there does not exist another solution y that *dominates* x, in the sense that y has at least the same profit and at most the same weight as x and at least one of these inequalities is strict. Generating the Pareto set is of great interest in many scenarios and widely used in practice. This approach fails to yield reasonable results in the worst case because even integer optimization problems with a simple combinatorial structure can have exponentially many Pareto-optimal solutions. In practice, however, generating the Pareto set is often feasible because typically the number of Pareto-optimal solutions does not attain its worst-case bound.

The discrepancy between practical experience and worst-case results motivates the study of the number of Pareto-optimal solutions in a more realistic scenario. One possible approach is to study the average number of Pareto-optimal solutions rather than the worst-case number. In order to analyze the average, one has to define a probability distribution on the set of instances with respect to which the average is taken. In most situations, however, it is not clear how to choose a probability distribution that reflects typical inputs. In order to bypass the limitations of worst-case and average-case analysis, Spielman and Teng [22] introduced the notion of *smoothed analysis*. They consider a semi-random input model in which first an adversary specifies an input that is afterwards slightly perturbed at random. The perturbation is motivated by the observation that in most practical applications instances are to some extent influenced by random events like, for example, measurement errors or numerical imprecision. Intuitively, the perturbation rules out pathological worst-case instances that are rarely observed in practice but dominate the worst-case analysis.

We consider bicriteria integer optimization problems in the framework of smoothed analysis. For this we assume that an adversary specifies an arbitrary set $\mathcal{S} \subseteq \mathcal{D}^n$ of solutions, where $\mathcal{D} \subseteq \mathbb{Z}$ denotes a finite set of integers, and two objective functions, profit $p \colon \mathcal{S} \to \mathbb{R}$ and weight $w \colon \mathcal{S} \to \mathbb{R}$. We assume that the profit is to be maximized while the weight is to be minimized. This assumption is without loss of generality as our results are not affected by changing the optimization direction of any of the objective functions. In our model, the weight function w can be chosen arbitrarily by the adversary, whereas the profit function p has to be linear of the form $p(x) = p_1 x_1 + \cdots + p_n x_n$.

In a classical worst-case analysis, the adversary can choose the coefficients $p_1, \ldots, p_n$ exactly so as to maximize the number of Pareto-optimal solutions. If he chooses these coefficients and the objective function w such that $p(x) = w(x)$ for all solutions $x \in \mathcal{S}$ and such that all solutions from $\mathcal{S}$ have pairwise different profits then all solutions from $\mathcal{S}$ are Pareto-optimal. In the model of smoothed analysis considered in this article, the adversary is less powerful: instead of being able to specify the coefficients exactly, he can only specify a probability distribution for each coefficient according to which it is chosen independently of the other coefficients. Allowing arbi-

trary distributions would include deterministic instances as a special case and hence, in order for the model to make sense, we need to ensure that the adversary is not too powerful and cannot concentrate the probability mass in a too small region. We achieve this by restricting the adversary to probability distributions that can be described by probability density functions that are bounded from above by some parameter $\phi \geq 1$. The parameter ϕ can be seen as a measure specifying how close the analysis is to a worst-case analysis: the larger ϕ, the more concentrated the probability mass can be and hence, the closer the analysis is to a worst-case analysis. To avoid that the effect of the perturbation is diminished by scaling, we assume that the distributions are normalized such that the expected absolute value of each profit is bounded from above by a constant.

To illustrate this model, let us consider the following example more reminiscent of Spielman and Teng's original model of smoothed analysis: first the adversary chooses an arbitrary vector of profits $(p_1, \ldots, p_n) \in [-1, 1]^n$, and then an independent Gaussian random variable with mean 0 and standard deviation $\sigma \leq 1$ is added to each profit p_i. In this example, the standard deviation σ takes over the role of ϕ: the smaller σ, the closer the analysis is to a worst-case analysis. For Gaussians with mean 0 and standard deviation σ, the maximum density is $1/(\sqrt{2\pi}\sigma) < 1/\sigma$ and the expected absolute value is $\sigma\sqrt{2/\pi} < \sigma \leq 1$. Hence, this model of Gaussian perturbations is covered by our model for $\phi = 1/\sigma$. In our model the adversary is even more powerful because he is not only allowed to choose the expected value of each profit p_i but also the type of noise as long as the density is bounded from above by ϕ. In particular, the adversary could choose for each profit p_i an interval $I_i \subseteq [0, 1]$ of length $1/\phi$ from which it is chosen independently uniformly at random.

The *smoothed number of Pareto-optimal solutions* is defined to be the maximum expected number of Pareto-optimal solutions for any choice of the set of solutions $\mathcal{S}$, the weight function w, and the distributions of the profits subject to the bounds on the maximum density and the expected absolute value. We present a new method for bounding the smoothed number of Pareto-optimal solutions, which yields an upper bound that is polynomial in the number n of variables, the parameter $\max_{a \in \mathcal{D}} |a|$, and the maximum density ϕ. This immediately implies polynomial upper bounds on the smoothed running time of several heuristics for generating the Pareto set of problems like the bounded knapsack problem. Previous results of this kind were restricted to the case of binary optimization problems. For this special case, our method yields an improved upper bound, which matches the known lower bound.

1.1 Related work

Multiobjective optimization is a well-studied research area. Several algorithms for generating the Pareto set of various optimization problems such as the (bounded) knapsack problem [11, 17], the bicriteria shortest path problem [6, 21], and the bicriteria network flow problem [9, 12] have been proposed. The running times of these algorithms depend crucially on the number of Pareto-optimal solutions and hence none of them 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 [13] study the number of Pareto-optimal solutions in multi-criteria shortest path problems experimentally. They consider networks 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. For more examples, we refer the reader to [8].

One way of coping with the bad worst-case behavior is to relax the requirement of finding the complete Pareto set. Papadimitriou and Yannakakis present a general framework for finding *approximate Pareto sets*. A solution x is ε-dominated by another solution y if $p(x)/p(y) \leq 1 + \varepsilon$ and $w(y)/w(x) \leq 1 + \varepsilon$. We say that $\mathcal{P}_\varepsilon$ is an ε-approximation of a Pareto set $\mathcal{P}$ if for any solution $x \in \mathcal{P}$ there is a solution $y \in \mathcal{P}_\varepsilon$ that ε-dominates it. Papadimitriou and Yannakakis [18] 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. Furthermore, they give a sufficient and necessary condition for the existence of an FPTAS for approximating the Pareto set of a multi-criteria optimization problem. Vassilvitskii and Yannakakis [24] and Diakonikolas and Yannakakis [7] investigate the problem of computing ε-approximate Pareto sets of small size.

After the seminal work of Spielman and Teng about the simplex algorithm [22], smoothed analysis has proven to be a good tool for narrowing the gap between practical experience and theoretical results, and it has been used to explain the practical success of algorithms in various areas. Some of these results are summarized in the surveys by Spielman and Teng [23] and and by Manthey and Röglin [15].

Beier and Vöcking [5] initiated the study of the smoothed number of Pareto-optimal solutions. They consider the special case of our model in which variables can take only binary values, i.e., $\mathcal{S} \subseteq \{0, 1\}^n$, and show that the smoothed number of Pareto-optimal solutions is bounded from above by $O(n^4\phi)$. Furthermore, they present a lower bound of $\Omega(n^2)$ on the expected number of Pareto-optimal solutions for profits that are chosen uniformly from the interval [0, 1]. Brunsch et al. [2] improve this lower bound to $\Omega(n^2\phi)$.

Röglin and Teng [20] study the same binary model, except that they allow any finite number d of perturbed linear objective functions (plus one arbitrary adversarial objective function) and that they require the densities according to which the coefficients are chosen to have a bounded support in $[-1, 1]$. They obtain an upper bound of $(n\phi)^{f(d)}$ for the smoothed number of Pareto-optimal solutions, where f grows exponentially in d. This result has been improved by Moitra and O'Donnell [14] to $O(n^{2d}\phi^{d(d+1)/2})$ and afterwards by Brunsch and Röglin [3] to $O(n^{2d}\phi^d)$ under the additional assumption that all densities are quasiconcave. Under this assumption they also prove an upper bound of $O((n^{2d}\phi^d)^c)$ for the c-th moment of the smoothed number of Pareto-optimal solutions for any constant $c \in \mathbb{N}$, which gives rise to non-trivial concentration bounds. Furthermore they also consider the case that $\mathcal{S} \subseteq \{0, 1, \dots, k\}^n$ and they analyze the effect of zero-preserving perturbations, in which the adversary can choose for each coefficient either a ϕ-bounded density function according to which it is chosen or set it deterministically to zero. The analyses in these articles are significantly more complicated than the upper bound presented in this article because they are targeted to problems with more than two objective functions.

Brunsch et al. [2] also prove a lower bound of $\Omega(n^{d-1.5}\phi^d)$ for the smoothed number of Pareto-optimal solutions in the case of d perturbed linear objective functions.

1.2 Model and notation

Let $\mathcal{D} \subseteq \mathbb{Z}$ be a finite set of integers and let $\mathcal{S} \subseteq \mathcal{D}^n$ denote an arbitrary set of solutions. We are interested in the cardinality of the set $\mathcal{P} \subseteq \mathcal{S}$ of Pareto-optimal solutions with respect to the objective functions profit $p\colon \mathcal{S} \to \mathbb{R}$ and weight $w\colon \mathcal{S} \to \mathbb{R}$, which are to be maximized and minimized, respectively. While the objective function w can be chosen arbitrarily by an adversary, the objective function p is assumed to be linear of the form $p(x) = p_1 x_1 + \ldots + p_n x_n$, where $x = (x_1, \ldots, x_n)^\top \in \mathcal{S}$. By abuse of notation, let p not only denote the objective function, but also the vector $(p_1, \ldots, p_n)^\top$. Then the profit $p(x)$ of a solution $x \in \mathcal{S}$ can be written as $p \cdot x$.

We assume that each p_i is a random variable that (independently of the other p_j) follows a density f_i with $f_i(x) \le \phi_i$ for all $x \in \mathbb{R}$. Furthermore, we denote by μ_i the expected absolute value of p_i, i.e., $\mu_i = \mathbf{E}[|p_i|] = \int_{x \in \mathbb{R}} |x| f_i(x)\, dx$. Let $\phi = \max_{i \in [n]} \phi_i$ and $\mu = \max_{i \in [n]} \mu_i$. For given parameters $\mathcal{D}, n, \phi$, and μ we assume that an adversary chooses the set $\mathcal{S}$, the objective function w, and the densities f_i so as to maximize the expected number of Pareto-optimal solutions. We refer to the largest expected number of Pareto-optimal solutions he can achieve as the *smoothed number of Pareto-optimal solutions* (with respect to $\mathcal{D}, n, \phi$, and μ).

We denote by $[n]$ the set $\{1, \ldots, n\}$, we use the notation $d = |\mathcal{D}|$ and $\Delta = \max\{a - b \mid a, b \in \mathcal{D}\}$, and we denote by H_n the nth harmonic number, i.e., $H_n = \sum_{i=1}^{n} 1/i$.

1.3 Our results

In this article, we present a new approach for bounding the smoothed number of Pareto-optimal solutions for bicriteria integer optimization problems. This approach follows to a large extent the analysis in [4] with one improvement from [19], which removes a factor of H_d. Altogether we obtain the following bound.

Theorem 1 *Let $\mathcal{D} \subseteq \mathbb{Z}$ be a finite set of integers, let $n \in \mathbb{N}$, and let $\mathcal{S} \subseteq \mathcal{D}^n$. Furthermore, let $w\colon \mathcal{S} \to \mathbb{R}$ be arbitrary and let arbitrary densities $f_1, \ldots, f_n$ be given according to which the coefficients $p_1, \ldots, p_n$ of the linear objective function p are chosen. Let ϕ_i denote an upper bound on f_i and let μ_i denote the expected absolute value of a random variable drawn according to f_i. Then the expected number of Pareto-optimal solutions in $\mathcal{S}$ with respect to p and w is at most*

$$4 \Delta d \left(\sum_{i=1}^{n} \phi_i \right) \left(\sum_{i=1}^{n} \mu_i \right) + dn + 1.$$

This implies that the smoothed number of Pareto-optimal solutions is bounded from above by $O(\Delta d \cdot n^2 \phi \mu)$. For $\mathcal{D} = \{0, \ldots, k\}$ and constant expected absolute value μ, the bound simplifies to $O(n^2 \phi \cdot k^2)$. For the binary case $\mathcal{D} = \{0, 1\}$ the bound further simplifies to $O(n^2 \phi)$. This improves significantly upon the previously known bound of $O(n^4 \phi)$ due to Beier and Vöcking [5]. The bound $O(n^2 \phi)$ follows also from the result of Moitra and O'Donnell [14] (which appeared after the conference version of this article [4]). However, the proof of Theorem 1 is much simpler than the analysis

of Moitra and O'Donnell, which is targeted to problems with more than two objective functions. It is also much simpler than the analysis of Beier and Vöcking [5] and it allows unbounded coefficients p_i (as long as the expected absolute value is bounded), whereas all results for more than two objectives assume that the coefficients have bounded support. Furthermore, all results for more than two objective functions in the literature either consider only the binary case (Moitra and O'Donnell [14]) or the dependence on k is much worse than k^2 (it is k^{32} in Brunsch and Röglin [3]). Hence none of these results implies Theorem 1.

Additionally, we also present two lower bounds on the smoothed number of Pareto-optimal solutions. In the following lower bound we use the term *ranking* to refer to the objective function w. The higher the ranking, the smaller is the weight.

Theorem 2 *Let $(5(c + 1) + 1) \log n \leq k \leq n^c$ for some $c \geq 2$ and assume that n is a multiple of $c + 2$. Let $\mathcal{D} = \{0, \ldots, k\}$ and $\mathcal{S} = \mathcal{D}^n$. There exists a ranking on $\mathcal{S}$ and a constant κ depending only on c such that the expected number of Pareto-optimal solutions is at least $\kappa n^2 k^2$ if each profit p_i is chosen independently uniformly at random from the interval $[-1, 1]$.*

In the proof of this lower bound we assume that the profits are chosen uniformly at random from the interval $[-1, 1]$ and hence, the lower bound holds for any ϕ with $\phi \geq 1$. The lower bound matches the upper bound in terms of n and k. We also obtain a slightly weaker lower bound for the case that the adversary is restricted to linear weight functions.

Theorem 3 *Let $\mathcal{D} = \{0, \ldots, k\}$ and $\mathcal{S} = \mathcal{D}^n$. Suppose that the profits are drawn independently at random according to a continuous probability distribution with non-increasing density function $f: \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$. Then there is a linear weight function $w: \mathcal{S} \to \mathbb{R}$ with coefficients $w_1, \ldots, w_n \in \mathbb{R}_{>0}$ for which the expected number of Pareto-optimal solutions is at least*

$$\frac{H_k}{4} k(n^2 - n) + kn + 1.$$

If the profits are drawn according to the uniform distribution over some interval $[0, a]$ with $a > 0$, then the above term equals the expected number of Pareto-optimal solutions.

While Theorem 2 shows that the upper bound in Theorem 1 is tight in terms of n and k, it is an open problem to find a lower bound that is additionally tight in terms of ϕ.

Knapsack Problem The *Nemhauser-Ullmann algorithm* solves the knapsack problem by enumerating the set of Pareto-optimal solutions [17]. This means that the capacity of the knapsack is neglected and all knapsack fillings that are Pareto-optimal with respect to profit and weight are enumerated. Then, the optimal solution of the knapsack problem is the Pareto-optimal solution with the highest weight not exceeding the capacity. The running time of this algorithm on an instance with n items is $\Theta(\sum_{i=1}^{n} q_i)$, where q_i denotes the number of Pareto-optimal solutions of the knapsack instance that consists only of the first i items.

We assume that an adversary can choose arbitrary weights and that he chooses a probability distribution for every profit arbitrarily subject to the bounds on the maximum density and the expected absolute value. Using their bound of $O(n^4\phi)$ on the smoothed number of Pareto-optimal solutions and linearity of expectation, Beier and Vöcking [5] show that the smoothed running time of the Nemhauser-Ullmann algorithm is bounded by $O(n^5\phi)$. Here the term smoothed running time refers to the maximum expected running time that can be achieved by choosing weights and probability distributions for the profits. Based on our improved bound on the expected number of Pareto-optimal solutions presented in Theorem 1, we conclude the following corollary.

Corollary 1 *The smoothed running time of the Nemhauser-Ullmann algorithm for the knapsack problem is $O(n^3\phi)$.*

For uniformly distributed profits Beier and Vöcking present a lower bound on the expected running time of $\Omega(n^3)$. Hence, we obtain tight bounds on the running time of the Nemhauser-Ullmann algorithm in terms of the number of items n.

Bounded Knapsack problem   In the bounded knapsack problem, a number $k \geq 2$ and a set of n items with weights and profits are given, and it is assumed that $k - 1$ copies of each of the n items are available. We assume that an adversary can choose arbitrary weights and that he chooses a probability distribution for every profit arbitrarily subject to the bounds on the maximum density and the expected absolute value. Then according to Theorem 1 the expected number of Pareto-optimal solutions is bounded from above by $O(n^2 k^2 \phi)$.

This observation alone does not yet imply that the Pareto set can be computed efficiently. However, Kellerer et al. [10] describe how an instance of the bounded knapsack problem with n items can be transformed into an instance of the knapsack problem with nK items, where $K = \Theta(\log k)$. We will call the items of this instance of the knapsack problem *virtual items* in the following. In the transformation, K numbers $\ell_1, \ldots, \ell_K \in \{0, 1, \ldots, k-1\}$ with $\sum_{i=1}^{K} \ell_i = k - 1$ are chosen and every item i in the bounded knapsack instance with profit p_i and weight w_i is replaced by K virtual items with profits $\ell_1 p_i, \ldots, \ell_K p_K$ and weights $\ell_1 w_i, \ldots, \ell_K w_K$. Using this transformation, the bounded knapsack problem can be solved by the Nemhauser-Ullmann algorithm in running time $\Theta(\sum_{i=1}^{nK} q_i)$, where q_i denotes the number of Pareto-optimal solutions of the knapsack instance that consists only of the first i virtual items.

In order to obtain an upper bound on the expected value of q_i, we can directly use Theorem 1. To see this, observe that the knapsack instance with only the first i virtual items can be viewed as an instance of the bounded knapsack problem in which only certain multiplicities of every item are allowed. Hence, for an appropriately chosen set S_i of solutions Theorem 1 applies and yields a bound of $O(n^2 k^2 \phi)$ for the expected value of q_i. Based on this, we obtain the following corollary.

Corollary 2 *The Nemhauser-Ullmann algorithm can be used to solve the bounded knapsack problem in smoothed running time $O(n^3 k^2 \log(k)\phi)$.*

Bicriteria shortest path problem   Different algorithms have been proposed for enumerating the Pareto set in bicriteria shortest path problems [6, 21]. An instance of the

bicriteria shortest path problem is described by a graph with n nodes and m edges in which each edge has certain costs and a certain length. The costs and the length of a path are then simply the sum of the costs and lengths of its edges, and the goal is to compute all Pareto-optimal paths. Given this Pareto set, one can in particular solve the constrained shortest path problem in which a budget is given and the goal is to find the shortest path whose costs do not exceed the given budget. As in the knapsack problem, the optimal solution to this problem is the Pareto-optimal solution with the largest costs not exceeding the budget.

Corley and Moon [6] suggest a modified version of the Bellman-Ford algorithm for enumerating the Pareto set of the bicriteria shortest path problem. Beier [1] shows that the running time of this algorithm is $O(nmU)$ where U is an upper bound on the number of Pareto-optimal solutions in certain sub-problems. These subproblems can be described by sets S of solutions that are subsets of $\{0, 1\}^m$. Given the bound on the smoothed number of Pareto-optimal solutions [5], Beier concludes that the smoothed running of this modified Bellman-Ford algorithm is $O(nm^5\phi)$ if either the costs or the lengths of the edges are perturbed. Based on Theorem 1, we obtain the following improved bound.

Corollary 3 *The smoothed running time of the modified Bellman-Ford algorithm is $O(nm^3\phi)$ if either the costs or the lengths of the edges are perturbed.*

In the following two sections, we prove the upper and lower bounds on the smoothed number of Pareto-optimal solutions.

2 Upper bound on the smoothed number of Pareto-optimal solutions

Since the profits are continuous random variables, the probability that there exist two solutions with exactly the same profit is zero. Hence, we can ignore this event and assume that no two solutions with the same profit exist. Furthermore, we assume without loss of generality that there are no two solutions with the same weight. If the adversary specifies a weight function in which two solutions have the same weight, we apply an arbitrary tie-breaking, which cannot decrease the expected number of Pareto-optimal solutions. We will now prove the upper bound on the smoothed number of Pareto-optimal solutions.

Proof We start the proof by defining d subsets of the Pareto set. We say that a Pareto-optimal solution x belongs to class $a \in \mathcal{D}$ if there exists an index $i \in [n]$ with $x_i \neq a$ such that the succeeding Pareto-optimal solution y satisfies $y_i = a$, where succeeding Pareto-optimal solution refers to the Pareto-optimal solution with the smallest weight among all solutions with higher profit than x (see Fig. 1). The Pareto-optimal solution with the highest profit, which does not have a succeeding Pareto-optimal solution, is not contained in any of the classes, but every other Pareto-optimal solution belongs to at least one of these classes. Let q denote the number of Pareto-optimal solutions and let q_a denote the number of Pareto-optimal solutions in class a. Since $q \leq 1 + \sum_{a \in \mathcal{D}} q_a$, linearity of expectation implies

Fig. 1 The solutions from $\mathcal{S}$ are depicted as points, where a solution $z \in \mathcal{S}$ corresponds to the point at $(w(z), p(z))$. Black points correspond to Pareto-optimal solutions and they dominate all solutions below the step function, which are depicted in gray. The Pareto-optimal solution y succeeds the Pareto-optimal solution x, which is class-a Pareto-optimal because there exists an index i with $x_i \neq a$ and $y_i = a$
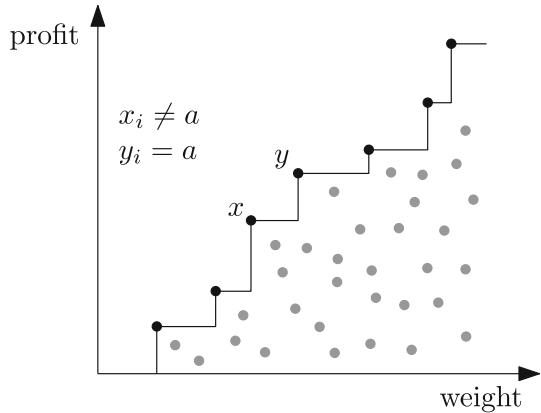
$$\mathbf{E}[q] \leq 1 + \sum_{a \in \mathcal{D}} \mathbf{E}[q_a]. \tag{1}$$

The following lemma, whose proof can be found below, shows an upper bound for the expected number of class-0 Pareto-optimal solutions.

Lemma 1 *The smoothed number of class-0 Pareto-optimal solutions is at most*

$$4\Delta \left(\sum_{i=1}^{n} \mu_i \right) \left(\sum_{i=1}^{n} \phi_i \right) + n.$$

To conclude the proof of the theorem, we show that counting the expected number of class-a Pareto-optimal solutions for $a \in \mathcal{D}$ with $a \neq 0$ can be reduced to counting the expected number of class-0 Pareto-optimal solutions.

Starting from the original set $\mathcal{S}$, we obtain a modified set $\mathcal{S}^a$ by subtracting the vector $(a, \ldots, a)$ from each solution vector $x \in \mathcal{S}$, that is, $\mathcal{S}^a = \{x - (a, \ldots, a) \mid x \in \mathcal{S}\}$. As $a \in \mathcal{D}$, the set $\mathcal{S}^a$ is a subset of $\mathcal{D}_a^n$, with $\mathcal{D}_a := \{x - a \mid x \in \mathcal{D}\} \subseteq \{-\Delta, -\Delta + 1, \ldots, \Delta - 1, \Delta\}$. This way, the profit of each solution x in $\mathcal{S}^a$ is smaller than the profit of its counterpart $x + (a, \ldots, a)$ in $\mathcal{S}$ by exactly $a \sum_{i=1}^{n} p_i$ if we extend the linear profit function p from $\mathcal{S}$ to $\mathcal{S}^a$. Let us additionally define a weight function $w^*: \mathcal{S}^a \to \mathbb{R}$ that assigns to every solution $x \in \mathcal{S}^a$ the weight that w assigns to its counterpart in $\mathcal{S}$.

Claim A solution in $\mathcal{S}^a$ is Pareto-optimal with respect to p and w^* if and only if its counterpart in $\mathcal{S}$ is Pareto-optimal with respect to p and w.

Proof Let x and y be two solutions from $\mathcal{S}$ and let x^a and y^a denote their counterparts in $\mathcal{S}^a$. Then $w(x) = w^*(x^a)$ and $w(y) = w^*(y^a)$. Furthermore $p(x) = p(x^a) + a \sum_{i=1}^{n} p_i$ and $p(y) = p(y^a) + a \sum_{i=1}^{n} p_i$. Hence, $p(x) > p(y)$ if and only if $p(x^a) > p(y^a)$. Overall this implies that x dominates y if and only if x^a dominates y^a. Since this is the case for every pair of solutions, the claim follows. □

A solution x is class-a Pareto-optimal in $\mathcal{S}$ if and only if the corresponding solution $x - (a, \ldots, a)$ is class-0 Pareto-optimal in $\mathcal{S}^a$. Hence, the number q_a of class-a Pareto-optimal solutions in $\mathcal{S}$ corresponds to the number $q_0(\mathcal{S}^a)$ of class-0 Pareto-optimal solutions in $\mathcal{S}^a$, which can be bounded by Lemma 1. Note that by definition of $\mathcal{D}_a$ we have $|\mathcal{D}_a| = |\mathcal{D}| = d$ as well as $\max\{b - c \mid b, c \in \mathcal{D}_a\} = \max\{b - c \mid b, c \in \mathcal{D}\} = \Delta$. Combining Equation (1) and Lemma 1 yields that $\mathbf{E}[q]$ is at most

$$1 + \sum_{a \in \mathcal{D}} \mathbf{E}\big[q_0(\mathcal{S}^a)\big] \leq 1 + \sum_{a \in \mathcal{D}} \left(4\Delta \left(\sum_{i=1}^{n} \mu_i \right) \left(\sum_{i=1}^{n} \phi_i \right) + n \right)$$

$$\leq 4\Delta d \left(\sum_{i=1}^{n} \mu_i \right) \left(\sum_{i=1}^{n} \phi_i \right) + dn + 1,$$

which proves the theorem. $\square$

To conclude the proof of Theorem 1, we prove Lemma 1.

Proof of Lemma 1 We assume $0 \in \mathcal{D}$ as otherwise there are no class-0 solutions. This assumption yields $|a| \leq \Delta$ for all $a \in \mathcal{D}$.

The main part of the proof is an upper bound on the probability that there exists a class-0 Pareto-optimal solution whose profit lies in a small interval $[t - \varepsilon, t)$, for some given $t \in \mathbb{R}$ and $\varepsilon > 0$. Roughly speaking, if ε is smaller than the smallest profit difference of any two Pareto-optimal solutions, then this probability equals the expected number of class-0 Pareto-optimal solutions in the interval $[t - \varepsilon, t)$. Then we can divide $\mathbb{R}$ into intervals of length ε and sum these expectations to obtain the desired bound on the expected number of Pareto-optimal solutions.

Let $t \in \mathbb{R}$ be chosen arbitrarily. We define x^* to be the solution from $\mathcal{S}$ with the lowest weight among all solutions satisfying the constraint $p \cdot x \geq t$, that is,

$$x^* = \underset{x \in \mathcal{S}: p \cdot x \geq t}{\operatorname{argmin}} \ w(x).$$

If there does not exist a solution $x \in \mathcal{S}$ with $p \cdot x \geq t$ then x^* does not exist. Otherwise, the solution x^* is Pareto-optimal. Let $\hat{x}$ denote the Pareto-optimal solution that precedes x^*, that is,

$$\hat{x} = \underset{x \in \mathcal{S}: w(x) < w(x^*)}{\operatorname{argmax}} \ p \cdot x.$$

See Fig. 2 for an illustration of these definitions. We aim at bounding the probability that $\hat{x}$ is a class-0 Pareto-optimal solution whose profit falls into the interval $[t - \varepsilon, t)$.

For this we classify class-0 Pareto-optimal solutions to be *ordinary* or *extraordinary*. Considering only ordinary solutions allows us to prove a bound that depends not only on the length ε of the interval but also on $|t|$, the distance to zero. This captures the intuition that it becomes increasingly unlikely to observe solutions whose profits are much larger than the expected profit of the most profitable solution. The final bound is

Fig. 2 Illustration of the definitions of x^* and $\hat{x}$. If $\hat{x}$ exists then it is Pareto-optimal. If additionally there is an index i with $x_i^* = 0$ and $\hat{x}_i \neq 0$ then $\hat{x}$ is a class-0 Pareto-optimal solution
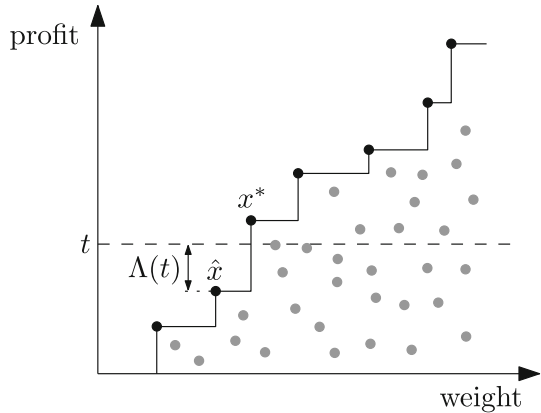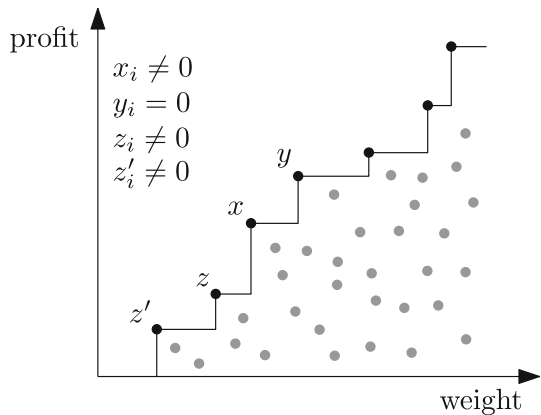

Fig. 3 The solution x is an extraordinary class-0 Pareto-optimal solution if $z_i \neq 0$ and $z_i' \neq 0$ holds for all indices $i \in [n]$ for which $x_i \neq 0$ and $y_i = 0$

obtained by observing that there can be at most n extraordinary class-0 Pareto-optimal solutions.

We would like to mention that the classification into ordinary and extraordinary solutions is only necessary because we allowed density functions with unbounded support for the p_i. If all densities would have a bounded support on, e.g., $[-1, 1]$ then the separate treatment of extraordinary solutions would not be necessary.

We classify solutions to be ordinary or extraordinary as follows. Let x be a class-0 Pareto-optimal solution and let y be the succeeding Pareto-optimal solution, which must exist by definition. We say that x is extraordinary if for all indices $i \in [n]$ with $x_i \neq 0$ and $y_i = 0$, *all* Pareto-optimal solutions z that precede x satisfy $z_i \neq 0$. In other words, for those indices i that make x class-0 Pareto-optimal, y is the Pareto-optimal solution with the smallest profit that is independent of p_i (see Fig. 3). We classify a class-0 Pareto-optimal solutions as ordinary if it is not extraordinary. For every index $i \in [n]$, there can be at most one extraordinary class-0 Pareto-optimal solution. In the following, we restrict ourselves to solutions $\hat{x}$ that are ordinary, and we denote by $\mathcal{P}^0$ the set of ordinary class-0 Pareto-optimal solutions. We define the *loser gap* to be the slack of the solution $\hat{x}$ from the threshold t (see Figure 2), that is,

$$\Lambda(t) = \begin{cases} t - p \cdot \hat{x} & \text{if } x^* \text{ and } \hat{x} \text{ exist and } \hat{x} \in \mathcal{P}^0, \\ \infty & \text{otherwise.} \end{cases}$$

If $\Lambda(t) \leq \varepsilon$, then there exists a solution $x \in \mathcal{P}^0$ with $p \cdot x \in [t - \varepsilon, t)$, namely $\hat{x}$. The converse is not true because it might be the case that $\hat{x} \notin \mathcal{P}^0$ and that there exists another solution $x \in \mathcal{P}^0$ with $p \cdot x \in [t - \varepsilon, t)$. If, however, ε is smaller than the minimum profit difference of any two Pareto-optimal solutions, then the existence of a solution $x \in \mathcal{P}^0$ with $p \cdot x \in [t - \varepsilon, t)$ implies $\hat{x} = x$ and hence $\Lambda(t) \leq \varepsilon$. Let $\mathcal{F}(\varepsilon)$ denote the event that there are two Pareto-optimal solutions whose profits differ by at most ε, then

$$\mathbf{Pr}\left[\exists x \in \mathcal{P}^0 : p \cdot x \in [t - \varepsilon, t) \,\middle|\, \neg\mathcal{F}(\varepsilon)\right] = \mathbf{Pr}\left[\Lambda(t) \leq \varepsilon \,\middle|\, \neg\mathcal{F}(\varepsilon)\right]. \quad (2)$$

In the following, we estimate, for a given $b > 0$, the expected number of Pareto-optimal solutions whose profits lie in the interval $(-b, b]$. For this, we partition the interval $(-b, b]$ into $2bm$ sub-intervals of length $1/m$ each, and we let the number $2bm$ of sub-intervals tend infinity. For $m \in \mathbb{N}$ and $i \in \{0, \ldots, 2bm - 1\}$, we set $I_i^m = (b_i, b_{i+1}]$ with $b_i = -b + i/m$. Since the number of Pareto-optimal solutions is always bounded by $|\mathcal{S}| \leq (d)^n$, we obtain

$$\mathbf{E}\left[|\mathcal{P}^0|\right] \leq \lim_{m \to \infty} \left(\mathbf{Pr}[\neg\mathcal{F}(1/m)] \cdot \mathbf{E}\left[|\mathcal{P}^0| \,\middle|\, \neg\mathcal{F}(1/m)\right] + \mathbf{Pr}[\mathcal{F}(1/m)] \cdot (d)^n \right).$$

The probability that two given solutions have a profit difference of at most ε can be bounded from above by $2\varepsilon\phi$. In order to see this, consider two solutions $x \neq y$ and choose an index i with $x_i \neq y_i$. Then use the principle of deferred decisions and assume that all p_j with $j \neq i$ are already fixed arbitrarily. Then the event $|p \cdot x - p \cdot y| \leq \varepsilon$ is equivalent to the event that p_i takes a value in a fixed interval (depending on the p_j with $j \neq i$) of length 2ε. Since the density of p_i is bounded from above by ϕ, the probability of the event $|p \cdot x - p \cdot y| \leq \varepsilon$ is at most $2\varepsilon\phi$. Hence, a union bound over all pairs of solutions (there are at most d^{2n} pairs) yields

$$\mathbf{Pr}[\mathcal{F}(1/m)] \leq 2d^{2n}\phi/m,$$

which tends to 0 when m tends to infinity. Hence, it holds

$$\mathbf{E}\left[|\mathcal{P}^0|\right] \leq \lim_{m \to \infty} \left(\mathbf{Pr}[\neg\mathcal{F}(1/m)] \cdot \mathbf{E}\left[|\mathcal{P}^0| \,\middle|\, \neg\mathcal{F}(1/m)\right] \right). \quad (3)$$

Under the condition $\neg\mathcal{F}(1/m)$, every interval I_i^m can contain at most one Pareto-optimal solution, and hence, under this condition, the probability that I_i^m contains a Pareto-optimal solution from $\mathcal{P}^0$ equals the expected number of Pareto-optimal solutions from $\mathcal{P}^0$ in I_i^m, yielding together with (2) and (3) that the expected number of ordinary class-0 Pareto-optimal solutions with profits in $(-b, b]$ is bounded from above by

$$\lim_{m\to\infty} \left(\mathbf{Pr}[\neg\mathcal{F}(1/m)] \cdot \sum_{i=0}^{2bm-1} \mathbf{Pr}\left[\exists x \in \mathcal{P}^0 : p \cdot x \in I_i^m \ \middle|\ \neg\mathcal{F}(1/m) \right] \right)$$

$$= \lim_{m\to\infty} \left(\mathbf{Pr}[\neg\mathcal{F}(1/m)] \cdot \sum_{i=0}^{2bm-1} \mathbf{Pr}\left[\Lambda(b_{i+1}) \le 1/m \mid \neg\mathcal{F}(1/m) \right] \right)$$

$$= \lim_{m\to\infty} \left(\sum_{i=0}^{2bm-1} \mathbf{Pr}\left[(\Lambda(b_{i+1}) \le 1/m) \wedge (\neg\mathcal{F}(1/m)) \right] \right)$$

$$\le \lim_{m\to\infty} \sum_{i=0}^{2bm-1} \mathbf{Pr}\left[\Lambda(b_{i+1}) \le 1/m \right]. \tag{4}$$

The only missing part is to analyze the probability of the event $\Lambda(t) \le \varepsilon$ for given $t \in \mathbb{R}$ and $\varepsilon > 0$, which is done in the following lemma.

Lemma 2 *For all $t \in \mathbb{R}$ and $\varepsilon > 0$,*

$$\mathbf{Pr}[\Lambda(t) \le \varepsilon] \le \mathbf{Pr}\left[\Delta \cdot \sum_{i=1}^{n} |p_i| \ge |t| \right] \cdot 2\varepsilon \sum_{i=1}^{n} \phi_i.$$

Lemma 2 yields the following upper bound on (4):

$$2\left(\sum_{i=1}^{n} \phi_i \right) \cdot \lim_{m\to\infty} \sum_{i=0}^{2bm-1} \left(\frac{1}{m} \cdot \mathbf{Pr}\left[\Delta \cdot \sum_{i=1}^{n} |p_i| \ge |b_i| \right] \right).$$

We consider $\mathbf{Pr}\left[\Delta \cdot \sum_{i=1}^{n} |p_i| \ge |t| \right]$ as a function of t. Because the density of p_i is bounded from above, this function is continuous. Therefore by the definition of the Riemann integral, we can rewrite the previous limit as

$$2\left(\sum_{i=1}^{n} \phi_i \right) \cdot \int_{-b}^{b} \mathbf{Pr}\left[\Delta \cdot \sum_{i=1}^{n} |p_i| \ge |t| \right] dt.$$

This term is an upper bound on the expected number of ordinary class-0 Pareto-optimal solutions in the interval $(-b, b]$. Letting b tend to infinity and using that the expected absolute value of profit p_i is μ_i yield that the expected number of ordinary class-0 Pareto-optimal solutions can be bounded from above by

$$2\left(\sum_{i=1}^{n} \phi_i \right) \cdot \int_{-\infty}^{\infty} \mathbf{Pr}\left[\Delta \cdot \sum_{i=1}^{n} |p_i| \ge |t| \right] dt$$

$$= 4\left(\sum_{i=1}^{n} \phi_i \right) \cdot \int_{0}^{\infty} \mathbf{Pr}\left[\Delta \cdot \sum_{i=1}^{n} |p_i| \ge t \right] dt$$

$$= 4 \left(\sum_{i=1}^{n} \phi_i \right) \cdot \mathbf{E} \left[\Delta \cdot \sum_{i=1}^{n} |p_i| \right]$$

$$= 4\Delta \left(\sum_{i=1}^{n} \mu_i \right) \left(\sum_{i=1}^{n} \phi_i \right).$$

Since there are at most n extraordinary class-0 Pareto-optimal solutions , this proves the lemma.   □

We conclude the proof of Theorem 1 and Lemma 1 by proving Lemma 2.

Proof of Lemma 2 In order to analyze the probability of the event $\Lambda(t) \leq \varepsilon$, we define a set of auxiliary random variables such that $\Lambda(t)$ is guaranteed to always take a value also taken by at least one of the auxiliary random variables. Then we analyze the auxiliary random variables and use a union bound to conclude the desired bound for $\Lambda(t)$.

Define $\mathcal{D}' = \mathcal{D} \setminus \{0\}$ and $\mathcal{S}^{x_i=v} = \{x \in \mathcal{S} \mid x_i = v\}$ for all $i \in [n]$ and $v \in \mathcal{D}$. We denote by $x^{*(i)}$ the solution from $\mathcal{S}^{x_i=0}$ with lowest weight with profit at least t, that is,

$$x^{*(i)} = \operatorname*{argmin}_{x \in \mathcal{S}^{x_i=0}:p\cdot x \geq t} w(x).$$

For each $i \in [n]$ we define the set $\mathcal{L}^i$ as follows. If there does not exist a solution $x \in \mathcal{S}^{x_i=0}$ with $p \cdot x \geq t$ then $x^{*(i)}$ does not exist. If $x^{*(i)}$ does exist and there also exists a solution in $\mathcal{S}^{x_i=0}$ with profit smaller than t, then $\mathcal{L}^i$ is defined as the set that consists of all solutions from $\bigcup_{v \in \mathcal{D}'} \mathcal{S}^{x_i=v}$ that have smaller weight than $x^{*(i)}$, otherwise $\mathcal{L}^i = \emptyset$. Let $\hat{x}^{(i)}$ denote the Pareto-optimal solution from the set $\mathcal{L}^i$ with the highest profit, that is,

$$\hat{x}^{(i)} = \operatorname*{argmax}_{x \in \mathcal{L}^i} p \cdot x.$$

Note that we must have $p \cdot \hat{x}^{(i)} < p \cdot x^{*(i)}$ because otherwise $p \cdot \hat{x}^{(i)} \geq p \cdot x^{*(i)} \geq t$ and $w \cdot \hat{x}^{(i)} < w \cdot x^{*(i)}$, contradicting the choice of $x^{*(i)}$. Finally, we define for each $i \in [n]$, the auxiliary random variable

$$\Lambda_i(t) = \begin{cases} t - p \cdot \hat{x}^{(i)} & \text{if } \hat{x}^{(i)} \text{ exists,} \\ \infty & \text{otherwise.} \end{cases}$$

Observe that the definitions of $\Lambda(t)$ and $\Lambda_i(t)$ are very similar. The only difference in the definitions of x^* and $x^{*(i)}$ is that we require $x_i^{*(i)} = 0$. The only difference in the definitions of $\hat{x}$ and $\hat{x}^{(i)}$ is that we require $\hat{x}_i^{(i)} \in \mathcal{D}'$. The reason for these additional constraints is that they will help us to apply the principle of deferred decisions. Intuitively, even if all p_j with $j \neq i$ are fixed arbitrarily, the randomness of p_i suffices to bound the probability of the event $\Lambda_i(t) \in (0, \varepsilon]$.

Lemma 3 *If $\Lambda(t) \leq \varepsilon$ then $\Lambda_i(t) \in (0, \varepsilon]$ for at least one $i \in [n]$.*

Proof Assume that $\Lambda(t) \leq \varepsilon$. Then by definition, x^* and $\hat{x}$ exist and $\hat{x} \in \mathcal{P}^0$, i.e., $\hat{x}$ is an ordinary class-0 Pareto-optimal solution. Since $\hat{x}$ is class-0 Pareto-optimal and x^* is the succeeding Pareto-optimal solution, there exists an index $i \in [n]$ such that

(a) $x_i^* = 0$ and $\hat{x}_i = v \neq 0$ for some $v \in \mathcal{D}'$, and
(b) there exists a solution $x \in \mathcal{S}^{x_i=0}$ with profit smaller than t.

The second condition is a consequence of the assumption that $\hat{x}$ is not extraordinary, that is, there exists a Pareto-optimal solution z with $z_i = 0$ that has smaller profit than $\hat{x}$ and hence smaller profit than t (this is important because otherwise by definition $\mathcal{L}^i = \emptyset$). Recall that $x^{*(i)}$ is defined to be the solution with the smallest weight in $\mathcal{S}^{x_i=0}$ with $p \cdot x \geq t$. As $x^* \in \mathcal{S}^{x_i=0}$, $x^* = x^{*(i)}$. Moreover, $\mathcal{L}^i$ consists of all solutions from $\bigcup_{v \in \mathcal{D}'} \mathcal{S}^{x_i=v}$ that have smaller weight than x^*. Thus, $\hat{x} \in \mathcal{L}^i$. By construction, $\hat{x}$ has the highest profit among the solutions in $\mathcal{L}^i$ and therefore, $\hat{x}^{(i)} = \hat{x}$ and $\Lambda_i(t) = \Lambda(t)$. □

We continue the proof by analyzing the probability of the event $\Lambda_i(t) \in (0, \varepsilon]$. If $\Lambda_i(t) \in (0, \varepsilon]$, which implies $\Lambda_i(t) \neq \infty$, then the following three events must occur simultaneously:

$\mathcal{E}_1$: There exists a solution $x \in \mathcal{S}^{x_i=0}$ with $p \cdot x \geq t$ (namely $x^{*(i)}$).
$\mathcal{E}_2$: There exists a solution $x \in \mathcal{S}^{x_i=0}$ with $p \cdot x < t$ (otherwise we defined $\mathcal{L}^i = \emptyset$).
$\mathcal{E}_3$: The solution $\hat{x}^{(i)}$ exists and its profit lies in the interval $[t - \varepsilon, t)$.

The events $\mathcal{E}_1$ and $\mathcal{E}_2$ depend only on the profits p_j, $j \neq i$. The existence and identity of $\hat{x}^{(i)}$ depends additionally on p_i, but the profits p_j, $j \neq i$, determine a set $\hat{X}^{(i)}$ of at most $d - 1$ candidate solutions such that $\hat{x}^{(i)} \in \hat{X}^{(i)}$ if $\hat{x}^{(i)}$ exists.

For each $i \in [n]$ and $v \in \mathcal{D}'$, we partition the set $\mathcal{L}^i = \bigcup_{v \in \mathcal{D}'} \mathcal{L}^{(i,v)}$ as follows. If $\mathcal{L}^i = \emptyset$, then we define $\mathcal{L}^{(i,v)} = \emptyset$. Otherwise $\mathcal{L}^{(i,v)}$ consists of all solutions from $\mathcal{L}^i \cap \mathcal{S}^{x_i=v}$, i.e., all solutions from $\mathcal{S}^{x_i=v}$ that have smaller weight than $x^{*(i)}$. Let $\hat{x}^{(i,v)}$ denote the Pareto-optimal solution from the set $\mathcal{L}^{(i,v)}$ with the highest profit, that is,

$$\hat{x}^{(i,v)} = \underset{x \in \mathcal{L}^{(i,v)}}{\arg\max}\, p \cdot x.$$

Let $\hat{X}^{(i)}$ denote the set that contains all $\hat{x}^{(i,v)}$ that exist (i.e., for which $\mathcal{L}^{(i,v)} \neq \emptyset$):

$$\hat{X}^{(i)} = \{\hat{x}^{(i,v)} \mid v \in \mathcal{D}' \wedge \hat{x}^{(i,v)} \text{ exists}\}.$$

If $\hat{x}^{(i)}$ exists then

$$\hat{x}^{(i)} = \arg\max\{p \cdot x \mid x \in \hat{X}^{(i)}\}.$$

For all $v \in \mathcal{D}'$ the existence and identity of $\hat{x}^{(i,v)}$ is completely determined by the profits p_j, $j \neq i$. Hence, if we fix all profits except for p_i, then $\hat{x}^{(i,v)}$ is fixed and its profit is $\kappa^{(i,v)} + v p_i$ for some constant

$$\kappa^{(i,v)} = p \cdot \hat{x}^{(i,v)} - v p_i = \sum_{j \neq i} p_j \hat{x}_j^{(i,v)}.$$

that depends only on the profits p_j with $j \neq i$.

With these definitions the event $\mathcal{E}_3$ is equivalent to the event

$\mathcal{E}_3'$: $\hat{X}^{(i)} \neq \emptyset$ and $\max \left\{ p \cdot x \mid x \in \hat{X}^{(i)} \right\}$ lies in the interval $[t - \varepsilon, t)$.

To analyze the probability that, given $\hat{X}^{(i)} \neq \emptyset$, $\max\{p \cdot x \mid x \in \hat{X}^{(i)}\}$ lies in the interval $[t - \varepsilon, t)$, we partition $\hat{X}^{(i)}$ into $\hat{X}^{(i, <0)}$ and $\hat{X}^{(i, >0)}$ with

$$\hat{X}^{(i, <0)} := \left\{ \hat{x}^{(i, v)} \mid v < 0 \text{ and } \hat{x}^{(i, v)} \in \hat{X}^{(i)} \right\}$$

and

$$\hat{X}^{(i, >0)} := \left\{ \hat{x}^{(i, v)} \mid v > 0 \text{ and } \hat{x}^{(i, v)} \in \hat{X}^{(i)} \right\}.$$

Then in order for $\mathcal{E}_3'$ to be true at least one of the following events must occur:

$\mathcal{E}_{(3, <0)}'$: $\hat{X}^{(i, <0)} \neq \emptyset$ and $\max \left\{ p \cdot x \mid x \in \hat{X}^{(i, <0)} \right\}$ lies in the interval $[t - \varepsilon, t)$,

$\mathcal{E}_{(3, >0)}'$: $\hat{X}^{(i, >0)} \neq \emptyset$ and $\max \left\{ p \cdot x \mid x \in \hat{X}^{(i, >0)} \right\}$ lies in the interval $[t - \varepsilon, t)$.

Given $\hat{X}^{(i, <0)} \neq \emptyset$, let

$$p_< = \max_{v \in \mathcal{D}'} \left\{ y \in \mathbb{R} \mid \hat{x}^{(i, v)} \in \hat{X}^{(i, <0)} \wedge \kappa^{(i, v)} + vy = t \right\}.$$

Claim Let $\hat{X}^{(i, <0)} \neq \emptyset$. For $p_i < p_<$ we have $p(\hat{x}^{(i, v)}) = \kappa^{(i, v)} + vp_i > t$ for at least one $\hat{x}^{(i, v)} \in \hat{X}^{(i, <0)}$.

Proof Consider an element $v \in \mathcal{D}'$ for which the maximum in the definition of $p_<$ is taken. For this v, we have $\kappa^{(i, v)} + vp_< = t$. Since $\hat{X}^{(i, <0)}$ contains only solutions $\hat{x}^{(i, v)}$ with $v < 0$, we obtain

$$p(\hat{x}^{(i, v)}) = \kappa^{(i, v)} + vp_i > \kappa^{(i, v)} + vp_< = t$$

for $p_i < p_<$. $\square$

We also have $\kappa^{(i, v)} + vp_< \leq t$ for all $v \in \mathcal{D}'$ with $\hat{x}^{(i, v)} \in \hat{X}^{(i, <0)}$. Therefore for $p_i > p_< + \varepsilon$ we obtain $\kappa^{(i, v)} + vp_i < \kappa^{(i, v)} + v(p_< + \varepsilon) \leq \kappa^{(i, v)} + v(p_<) - \varepsilon \leq t - \varepsilon$. In order for $\mathcal{E}_{(3, <0)}'$ to be true, we therefore must have $p_i \in [p_<, p_< + \varepsilon]$.

Given $\hat{X}^{(i, >0)} \neq \emptyset$, we analogously let

$$p_> = \min_{v \in \mathcal{D}'} \{ y \in \mathbb{R} \mid \hat{x}^{(i, v)} \in \hat{X}^{(i, >0)} \wedge \kappa^{(i, v)} + vy = t - \varepsilon \}.$$

Then for $p_i < p_>$ we have $p(\hat{x}^{(i, v)}) < t - \varepsilon$ for all $\hat{x}^{(i, v)} \in \hat{X}^{(i, >0)}$.

Claim For $p_i > p_> + \varepsilon$ we have $p(\hat{x}^{(i, v)}) > t$ for at least one $\hat{x}^{(i, v)} \in \hat{X}^{(i, >0)}$.

Proof Consider an element $v \in \mathcal{D}'$ for which the minimum in the definition of $p_>$ is taken. For this v, we have $\kappa^{(i,v)} + vp_> = t - \varepsilon$. Since $\hat{X}^{(i,>0)}$ contains only solutions $\hat{x}^{(i,v)}$ with $v > 0$, we obtain

$$p(\hat{x}^{(i,v)}) = \kappa^{(i,v)} + vp_i > \kappa^{(i,v)} + v(p_> + \varepsilon) = t - \varepsilon + v\varepsilon \geq t$$

for $p_i > p_> + \varepsilon$. $\qquad\square$

In order for $\mathcal{E}'_{(3,>0)}$ to be true we therefore must have $p_i \in [p_>, p_> + \varepsilon]$.

Note that $p_<$ and $p_>$ depend only on the profits already fixed, i.e., the profits p_j with $j \neq i$. Since the events $\mathcal{E}_1$ and $\mathcal{E}_2$ are independent of p_i, we obtain[1]

$$\begin{aligned}
\mathbf{Pr}[\mathcal{E}_3 \mid \mathcal{E}_1 \text{ and } \mathcal{E}_2] &= \mathbf{Pr}\left[\mathcal{E}'_3 \mid \mathcal{E}_1 \text{ and } \mathcal{E}_2\right] \\
&= \mathbf{Pr}\left[\max\{p \cdot x \mid x \in \hat{X}^{(i)}\} \in [t - \varepsilon, t) \mid \mathcal{E}_1 \text{ and } \mathcal{E}_2\right] \\
&\leq \mathbf{Pr}[\, p_i \in [p_<, p_< + \varepsilon] \cup [p_>, p_> + \varepsilon] \mid \mathcal{E}_1 \text{ and } \mathcal{E}_2] \\
&\leq 2\varepsilon\phi.
\end{aligned}$$

The event $\mathcal{E}_2$ implies $-\Delta \sum_{j=1}^{n} |p_j| < t$. This is equivalent to $\Delta \sum_{j=1}^{n} |p_j| \geq -t$ because the probability that the inequality is satisfied with equality is zero. For $t \leq 0$, the event $\mathcal{E}_1$ implies $\Delta \sum_{j=1}^{n} |p_j| \geq t$, and hence, for every $t \in \mathbb{R}$, one of the events implies $\Delta \sum_{j=1}^{n} |p_j| \geq |t|$. This yields

$$\mathbf{Pr}[\mathcal{E}_1 \text{ and } \mathcal{E}_2] \leq \mathbf{Pr}\left[\Delta \sum_{j=1}^{n} |p_j| \geq |t|\right].$$

Since the events $\mathcal{E}_1$ and $\mathcal{E}_2$ do not depend on p_i and in our analysis of $\mathcal{E}_3$ we assumed that all p_j with $j \neq i$ can be arbitrarily fixed, we obtain

$$\begin{aligned}
\mathbf{Pr}[\Lambda_i(t) \in (0, \varepsilon]] &\leq \mathbf{Pr}[\mathcal{E}_1 \text{ and } \mathcal{E}_2 \text{ and } \mathcal{E}_3] \\
&= \mathbf{Pr}[\mathcal{E}_1 \text{ and } \mathcal{E}_2] \cdot \mathbf{Pr}[\mathcal{E}_3 \mid \mathcal{E}_1 \text{ and } \mathcal{E}_2] \\
&\leq \mathbf{Pr}\left[\Delta \sum_{j=1}^{n} |p_j| \geq |t|\right] \cdot 2\varepsilon\phi_i.
\end{aligned}$$

To conclude the proof, we apply a union bound and Lemma 3:

$$\begin{aligned}
\mathbf{Pr}[\Lambda(t) \leq \varepsilon] &\leq \mathbf{Pr}[\exists i : \Lambda_i(t) \in (0, \varepsilon]] \\
&\leq \sum_{i=1}^{n} \mathbf{Pr}[\Lambda_i(t) \in (0, \varepsilon]]
\end{aligned}$$

---

[1] The contribution of [19] is this improved upper bound for $\mathbf{Pr}[\mathcal{E}_3]$. In [4], we used a union bound over all candidates from $\hat{X}^{(i)}$, which resulted in an additional factor of $O(\log k)$.

$$\leq \sum_{i=1}^{n} \mathbf{Pr}\left[\Delta \sum_{j=1}^{n} |p_j| \geq |t|\right] 2\varepsilon\phi_i$$

$$\leq \mathbf{Pr}\left[\Delta \cdot \sum_{i=1}^{n} |p_i| \geq |t|\right] \cdot 2\varepsilon \sum_{i=1}^{n} \phi_i.$$

$\square$

3 Lower bounds on the smoothed number of Pareto-optimal solutions

In this section we first present a lower bound of $\Omega(n^2 k \log k)$ on the smoothed number of Pareto-optimal solutions for $\mathcal{S} = \{0, \ldots, k\}^n$, generalizing a bound for the binary domain presented in [5]. Afterwards we prove a stronger bound of $\Omega(n^2 k^2)$ under stronger assumptions. The weaker bound provides a vector of weights $w_1, \ldots, w_n$ such that the bound holds for a linear weight function $w \cdot x$. For the stronger bound we can only prove that there is some weight function $w \colon \mathcal{S} \to \mathbb{R}$ for which the bound holds but this function might not be linear.

3.1 Lower bound for linear weight functions

For linear weight functions, we prove the following lower bound on the expected number of Pareto-optimal solutions restated in Theorem 3.

Theorem 3 *Let $\mathcal{D} = \{0, \ldots, k\}$ and $\mathcal{S} = \mathcal{D}^n$. Suppose that the profits are drawn independently at random according to a continuous probability distribution with non-increasing density function $f \colon \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$. Then there is a linear weight function $w \colon \mathcal{S} \to \mathbb{R}$ with coefficients $w_1, \ldots, w_n \in \mathbb{R}_{>0}$ for which the expected number of Pareto-optimal solutions is at least*

$$\frac{H_k}{4} k(n^2 - n) + kn + 1.$$

If the profits are drawn according to the uniform distribution over some interval $[0, a]$ with $a > 0$, then the above term equals the expected number of Pareto-optimal solutions.

Similarly, a lower bound of $\Omega(n^2 k \log k)$ can be obtained for the case that f is the density of a Gaussian random variable with mean 0. Since all weights w_i are larger than 0, a solution with a negative profit cannot be contained in a Pareto-optimal solution. Hence, we can ignore those items. Restricted to the interval $[0, \infty)$ the density of a Gaussian random variable with mean 0 is non-increasing and hence we can apply Theorem 3 when taking into account that with high probability at least a constant fraction of the random variables take positive values.

Proof of Theorem 3 The set $\mathcal{S} = \{0, \ldots, k\}^n$ corresponds to the solution set of the bounded knapsack problem in which up to k identical copies of each item can be put into the knapsack. For the sake of a simple presentation, we describe our construction in terms of this knapsack problem. We fix the weights of all items by setting $w_i = (k+1)^i$ for all $i \in [n]$. This way, the lexicographic order of the solutions in $\mathcal{S}$ is the same as the order defined by the weight $w \cdot x$ of solutions. Since the density function of the profits is assumed to be non-increasing, the distribution function $F \colon \mathbb{R}_{\geq 0} \to [0, 1]$ is concave as $F' = f$. Furthermore, $F(0) = 0$. Observe that such a function is sub-additive, that is, $F(a + b) \leq F(a) + F(b)$ for every $a, b \geq 0$.

Let $\mathcal{S}_j$ denote the set of the first $(k + 1)^j$ solutions in the lexicographic order, which are exactly those solutions that contain only copies of the items $1, \ldots, j$. We define $P_j = k \sum_{i=1}^{j} p_i$ and we denote by $\mathcal{P}_j$ the set of Pareto-optimal solutions over $\mathcal{S}_j$. Observe that the last solution in $\mathcal{S}_j$ has profit P_j and it is Pareto-optimal with probability 1.

For any given $\alpha > 0$, let X_α^j denote the number of Pareto-optimal solutions in $\mathcal{P}_j$ with profit at least $P_j - \alpha$, not counting the last solution in this sequence, which is $(k, \ldots, k, 0, \ldots, 0)$. By induction we show $\mathbf{E}\left[X_\alpha^j\right] \geq j \sum_{i=1}^{k} F(\alpha/i)$, where F denotes the distribution function of the profits. We partition the interval $[0, \infty)$ into disjoint intervals $I_0 = (\alpha, \infty), I_\ell = (\alpha/(\ell+1), \alpha/\ell]$ for $\ell \in [k-1]$, and $I_k = [0, \alpha/k]$. For every $i \in [n]$ and for $\ell \in \{0, \ldots, k\}$, we denote by A_ℓ^i the event that p_i lies in the interval I_ℓ. For all $\ell \in [k-1]$ it holds $\mathbf{Pr}\left[A_\ell^i\right] = F(\alpha/\ell) - F(\alpha/(\ell+1))$. Furthermore we have $\mathbf{Pr}\left[A_0^i\right] = 1 - F(\alpha)$ and $\mathbf{Pr}\left[A_k^i\right] = F(\alpha/k)$.

Claim For $j = 1$, the base case of the induction, we have

$$\mathbf{E}\left[X_\alpha^1\right] = \sum_{\ell=1}^{k} F\left(\frac{\alpha}{\ell}\right).$$

Proof Since $w_1 > 0$ and $p_1 > 0$ with probability one, we have $\mathcal{P}_1 = \mathcal{S}_1 = \{(i, 0, \ldots, 0) \mid i \in \{0, \ldots, k\}\}$. By definition X_α^1 counts the number of solutions in $\mathcal{P}_1$ with profit at least $P_1 - \alpha = kp_1 - \alpha$, not counting the last solution $(k, 0, \ldots, 0)$. Hence,

$$X_\alpha^1 = \{i \in \{0, \ldots, k-1\} \mid ip_1 \geq kp_1 - \alpha\} = \{i \in \{0, \ldots, k-1\} \mid k - i \leq \alpha/p_1\}$$
$$= \min\{\lfloor \alpha/p_1 \rfloor, k\}.$$

This implies that $X_\alpha^1 = \ell$ holds if and only if the event A_ℓ^1 occurs. Hence,

$$\mathbf{E}\left[X_\alpha^1\right] = \sum_{\ell=0}^{k} \ell \cdot \mathbf{Pr}\left[A_\ell^1\right] = k \cdot \mathbf{Pr}\left[A_k^1\right] + \sum_{\ell=1}^{k-1} \ell\left(F\left(\frac{\alpha}{\ell}\right) - F\left(\frac{\alpha}{(\ell+1)}\right)\right)$$
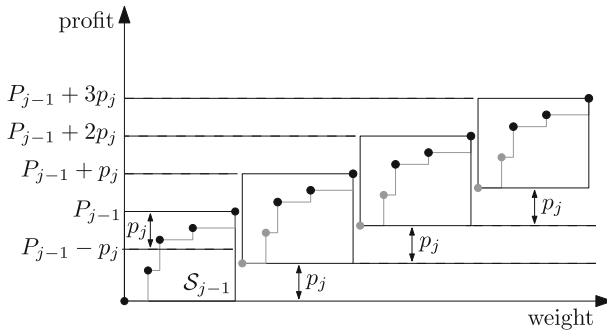$$= \sum_{\ell=1}^{k} F\left(\frac{\alpha}{\ell}\right).$$

Fig. 4 In this example, we have $k = 3$ and hence the solutions in $\mathcal{S}_j$ are grouped into 4 blocks, with block $\ell \in \{0, \ldots, 3\}$ containing all solutions with $x_j = \ell$. These blocks are depicted as rectangles. Block 0 (the leftmost block) corresponds to $\mathcal{S}_{j-1}$ and the other blocks are shifted copies of block 0. The points inside the blocks depict the Pareto-optimal solutions from $\mathcal{P}_{j-1}$. Let $\ell \geq 1$. Our choice of the weights guarantees that each solution in block ℓ has more weight than any solution from a block $\ell' < \ell$ and hence solutions from block ℓ cannot dominate solutions from a block $\ell' < \ell$. On the other hand any solution from a block $\ell' < \ell$ has profit at most $P_{j-1} + (\ell - 1)p_j$ and hence solutions from block ℓ with a larger profit cannot be dominated by solutions from a block $\ell' < \ell$. Since every block is a shifted copy of block 0, these solutions correspond to solutions from block 0 with a profit larger than $P_{j-1} - p_j$ (there are three such solutions in the example). Hence, exactly these solutions give rise to one new Pareto-optimal solution in each of the k following blocks. Pareto-optimal solutions in $\mathcal{P}_j$ are marked in black

□

Now we consider the case $j > 1$. We group the solutions in $\mathcal{S}_j$ into $k + 1$ blocks, with block $\ell \in \{0, \ldots, k\}$ containing all solutions with $x_j = \ell$. Block 0 corresponds to $\mathcal{S}_{j-1}$. Each Pareto-optimal solution in $\mathcal{S}_{j-1}$ with profit in the interval $(P_{j-1} - p_j, P_{j-1}]$ gives rise to one new Pareto-optimal solution in each of the k following blocks (see Fig. 4). In the event A_0^j we have $X_\alpha^j = X_\alpha^{j-1}$ because all solutions that contribute to X_α^j are in block k. In the event A_1^j we have $X_{p_j}^{j-1} + 1$ Pareto-optimal solutions in block k and $X_{\alpha - p_j}^{j-1} + 1$ Pareto-optimal solutions in block $k - 1$. Since the last solution is not counted in X_α^j, we have $X_\alpha^j = X_{p_j}^{j-1} + X_{\alpha - p_j}^{j-1} + 1$ (see Fig. 5). By similar reasoning, event A_ℓ^j implies $X_\alpha^j = \ell X_{p_j}^{j-1} + X_{\alpha - lp_j}^{j-1} + \ell$. Hence, it follows that we can lower bound the expected value of X_α^j by

$$\sum_{\ell=0}^{k} \mathbf{Pr}\left[A_\ell^j\right] \left(\ell \cdot \mathbf{E}\left[X_{p_j}^{j-1} \mid A_\ell^j\right] + \mathbf{E}\left[X_{\alpha - \ell p_j}^{j-1} \mid A_\ell^j\right] + \ell\right)$$

$$= \sum_{\ell=0}^{k} \int_{x \in I_\ell} f(x) \cdot \left(\ell \cdot \mathbf{E}\left[X_x^{j-1}\right] + \mathbf{E}\left[X_{\alpha - \ell x}^{j-1}\right] + \ell\right) dx$$

$$\geq \sum_{\ell=0}^{k} \int_{x \in I_\ell} f(x) \cdot \left(\ell \cdot (j-1) \sum_{i=1}^{k} F\left(\frac{x}{i}\right) + (j-1) \sum_{i=1}^{k} F\left(\frac{\alpha - lx}{i}\right) + \ell\right) dx,$$

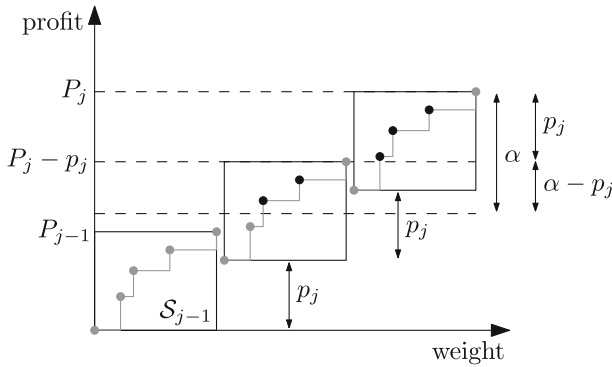where the last inequality follows from the induction hypothesis. We can further rewrite this term as

Fig. 5 In this example, we have $k = 2$ and hence the solutions in $\mathcal{S}_j$ are grouped into 3 blocks, with block $\ell \in \{0, \ldots, 2\}$ containing all solutions with $x_j = \ell$. These blocks are depicted as rectangles. Block 0 (the leftmost block) corresponds to $\mathcal{S}_{j-1}$ and the other blocks are shifted copies of block 0. The points inside the blocks depict the Pareto-optimal solutions from $\mathcal{P}_{j-1}$. The Pareto-optimal solutions from $\mathcal{P}_{j-1}$ in block 0 and their shifted copies in the other blocks together constitute a superset of the Pareto set $\mathcal{P}_j$. In this example, event A_1^j occurs, i.e., $p_j \in (\alpha/2, \alpha]$. Hence, for X_α^j only the last two blocks are relevant. It holds $X_\alpha^j = X_{p_j}^{j-1} + X_{\alpha-p_j}^{j-1} + 1$. The term $X_{p_j}^{j-1}$ counts the number of relevant Pareto-optimal solutions in block 2 (the rightmost block). It holds $X_{p_j}^{j-1} = 3$ and the corresponding solutions are marked in black. The term $X_{\alpha-p_j}^{j-1}$ counts the number of relevant solutions in block 1. It holds $X_{\alpha-p_j}^{j-1} = 2$ and the corresponding solutions are marked in black. Observe that the last solution in block 1 with profit $P_j - p_j$ needs to be counted in X_α^j. However, this solution is not counted in $X_{j-1}^{\alpha-p_j}$ by definition. This is the reason for the $+1$ term in the formula for X_α^j above

$$\sum_{\ell=0}^{k} \int_{x \in I_\ell} f(x) \cdot \left((j-1) \sum_{i=1}^{k} \left[\ell \cdot F\left(\frac{x}{i}\right) + F\left(\frac{\alpha - \ell x}{i}\right) \right] + \ell \right) dx$$

$$\geq \sum_{\ell=0}^{k} \int_{x \in I_\ell} f(x) \cdot \left((j-1) \sum_{i=1}^{k} F\left(\frac{\alpha}{i}\right) + \ell \right) dx$$

$$= (j-1) \sum_{i=1}^{k} F\left(\frac{\alpha}{i}\right) + \sum_{\ell=0}^{k} \ell \cdot \mathbf{Pr}\left[A_\ell^j\right] = j \sum_{i=1}^{k} F\left(\frac{\alpha}{i}\right),$$

where the inequality is due to the fact that the function F is sub-additive. If every profit is chosen uniformly at random from some interval $[0, a]$ with $a > 0$, then this term equals exactly the expected number of Pareto-optimal solutions.

Now let $Y_j = |\mathcal{P}_j| - |\mathcal{P}_{j-1}|$ denote the number of new Pareto-optimal solutions in $\mathcal{P}_j$. Observe that $Y_j = kX_{p_j}^{j-1} + k$. This follows from the fact that each Pareto-optimal solution in $\mathcal{P}_{j-1}$ with profit in the interval $(P_{j-1} - p_j, P_{j-1}]$ gives rise to k new Pareto-optimal solutions (see Fig. 4). The additive k is due to the fact that the last solution in $\mathcal{P}_{j-1}$ is not counted in $X_{p_j}^{j-1}$ but yields k new solutions in $\mathcal{P}_j$. Since p_j and X_α^{j-1} are independent, the induction hypothesis implies

$$\mathbf{E}\big[Y_j\big] = \mathbf{E}\Big[kX_{p_j}^{j-1} + k\Big] \geq \mathbf{E}\Bigg[k(j-1)\sum_{i=1}^{k} F\left(\frac{p_j}{i}\right) + k\Bigg].$$

Furthermore, the number of Pareto-optimal solutions in $\mathcal{P}_n$ is $q = 1 + \sum_{j=1}^{n} Y_j$. The additional 1 is due to the first solution $(0, \ldots, 0)$, which is always Pareto-optimal. Therefore,

$$\mathbf{E}[q] = 1 + \sum_{j=1}^{n} \mathbf{E}\big[Y_j\big] = 1 + \sum_{j=1}^{n} \mathbf{E}\Big[kX_{p_j}^{j-1} + k\Big]$$

$$\geq 1 + \sum_{j=1}^{n} \mathbf{E}\Bigg[k(j-1)\sum_{i=1}^{k} F\left(\frac{p_j}{i}\right) + k\Bigg].$$

The random variable $F(p_j)$ is uniformly distributed over the interval $[0, 1]$. To see this, observe that for any $\alpha \in [0, 1]$ we have

$$\mathbf{Pr}\big[F(p_j) \leq \alpha\big] = \mathbf{Pr}\Big[p_j \leq F^{-1}(\alpha)\Big] = F(F^{-1}(\alpha)) = \alpha,$$

where F^{-1} denotes the inverse function of F. This function is not unique in general because the distribution F is not injective in general. However, the argument works for any choice of F^{-1}. Thus $\mathbf{E}\big[F(p_j)\big] = 1/2$. As F is sub-additive, $i \cdot F(p_j/i) \geq F(p_j)$ holds, which implies $\mathbf{E}\big[F(p_j/i)\big] \geq \mathbf{E}\big[F(p_j)/i\big] = 1/(2i)$. Using $\mathbf{E}\Big[\sum_{i=1}^{k} F(p_j/i)\Big] = \frac{1}{2} H_k$ yields

$$\mathbf{E}[q] \geq \frac{H_k}{4} k(n^2 - n) + kn + 1.$$

If the profits are drawn according to the uniform distribution over some interval $[0, a]$ with $a > 0$, then the above inequality holds with equality. $\qquad \square$

3.2 Lower bound for general weight functions

Every weight function induces a *ranking* on the set of solutions, and in the following, we use the terms weight function and ranking synonymously. We assume that k is a function of n with $(5(c+1)+1)\log n \leq k \leq n^c$ for some constant c. We use the *probabilistic method* to show that, for each sufficiently large $n \in \mathbb{N}$, a ranking exists for which the expected number of Pareto-optimal solutions is lower bounded by $\Omega(n^2 k^2)$. That is, we create a ranking at random (but independently of the profits) and show that the expected number of Pareto-optimal solutions (where the expectation is taken over both the random ranking and the random profits) satisfies the desired lower bound. This implies that, for each sufficiently large $n \in \mathbb{N}$, there must exist a deterministic ranking on $\{0, \ldots, k\}^n$ for which the expected number of Pareto-optimal solutions (where the expectation is now taken only over the random profits) is $\Omega(n^2 k^2)$.

Theorem 2 *Let $(5(c + 1) + 1) \log n \leq k \leq n^c$ for some $c \geq 2$ and assume that n is a multiple of $c + 2$. Let $\mathcal{D} = \{0, \ldots, k\}$ and $\mathcal{S} = \mathcal{D}^n$. There exists a ranking on $\mathcal{S}$ and a constant κ depending only on c such that the expected number of Pareto-optimal solutions is at least $\kappa n^2 k^2$ if each profit p_i is chosen independently uniformly at random from the interval $[-1, 1]$.*

Before we describe how the ranking is created, we want to give a short overview on the ideas of the proof. In order to show a lower bound of $\Omega(n^2 k^2)$ on the expected number of Pareto-optimal solutions for $\mathcal{S} = \{0, \ldots, k\}^n$, we will use a similar approach as in Sect. 3.1. However, in order to obtain the higher bound, we need a larger number of items. We will use the existing original items to create new items called *virtual items*. As virtual items we will allow specific subsets of the n original items. We will randomly choose some of the virtual items and again similar to Sect. 3.1 we will create a ranking which can be represented as a linear function on binary combinations of the virtual items. Note that we might not be able to represent such a ranking by a linear function on the original items. The proof that this creates an expected number of at least $\Omega(n^2 k^2)$ Pareto-optimal solutions consists of two parts. In the first part we show that it is likely that the random set of virtual items creates a feasible instance and that, in case the profits of the original items are chosen uniformly at random, the distribution of the profits of the possible virtual items is likely to be close to a uniform distribution. In the second part we then show how to apply a similar proof as for Theorem 3 for a set of randomly chosen virtual items.

In order to describe how the ranking is created, we define *virtual items*. Let $[n]$ be the set of *original items* and assume that we have k instances of each of these n items. A virtual item is a vector $x \in \mathcal{D}^n$. Intuitively, adding the virtual item x to the knapsack corresponds to inserting x_i instances of the ith original item into the knapsack for every $i \in [n]$.

Assume that a sequence $x^{(1)}, \ldots, x^{(\ell)}$ of virtual items is given. Based on this sequence, we create a ranking on the set of solutions $\mathcal{D}^n$ similar to the ranking used in Theorem 3 but for the binary case in which every virtual item can be "contained" at most once in every solution. That is, we create a ranking such that solutions that "contain" the ith virtual item cannot dominate solutions that "consist" only of a subset of the first $i - 1$ virtual items. Let $\mathcal{S}_0 = \{(0, \ldots, 0)\}$ and assume that the solution $(0, \ldots, 0)$ has the highest rank, i.e., that it cannot be dominated by any other solution. Let $\mathcal{S}_i$ denote the set of solutions that can be obtained by adding a subset of the first i virtual items, that is,

$$\mathcal{S}_i = \mathcal{S}_{i-1} \cup \left\{ x + x^{(i)} \mid x \in \mathcal{S}_{i-1} \right\}.$$

Let $\mathcal{S}_i^* = \mathcal{S}_i \setminus \mathcal{S}_{i-1}$. In the ranking we define, each solution from $\mathcal{S}_i^*$ is ranked lower than every solution from $\mathcal{S}_{i-1}$. It remains to define the ranking among two solutions $x, y \in \mathcal{S}_i^*$. The solutions x and y can uniquely be written as $x = x' + x^{(i)}$ and $y = y' + x^{(i)}$ for some $x', y' \in \mathcal{S}_{i-1}$. Based on this observation, we define the ranking between x and y to be the same as the one between x' and y'. Furthermore, we define the ranking in such a way that all solutions in $\mathcal{S} \setminus \mathcal{S}_\ell$ are ranked lower than all solutions

in $\mathcal{S}_\ell$. Hence, we do not need to consider the solutions in $\mathcal{S} \setminus \mathcal{S}_\ell$ anymore. For a given sequence of virtual items, this yields a fixed ranking among the solutions in $\mathcal{S}_\ell$.

Example 1 In order to illustrate the way of how the ranking is created, let us give an example. Let us assume that $n = 3$ and that three virtual items are chosen, namely $x^{(1)} = (1, 0, 1)$, $x^{(2)} = (1, 1, 0)$, and $x^{(3)} = (0, 0, 1)$. Then $\mathcal{S}_0 = \{(0, 0, 0)\}$, $\mathcal{S}_1 = \{(0, 0, 0), (1, 0, 1)\}$, $\mathcal{S}_2 = \{(0, 0, 0), (1, 0, 1), (1, 1, 0), (2, 1, 1)\}$, and $\mathcal{S}_3 = \{(0, 0, 0), (1, 0, 1), (1, 1, 0), (2, 1, 1), (0, 0, 1), \quad (1, 0, 2), (1, 1, 1), (2, 1, 2)\}$. The solutions in $\mathcal{S}_3$ are listed according to the ranking, that is, $(0, 0, 0)$ is the highest ranked solution and $(2, 1, 2)$ is the lowest ranked solution. $\qquad\qquad\Box$

Now we describe how the sequence of virtual items is chosen. We set $\ell :=$ $nk/(2e(c + 2))$. Since we assumed that n is a multiple of $c + 2$, we can partition the set of original items into $c + 2$ groups with $n' = n/(c + 2)$ items each. Let V denote the set of virtual items that contain one item from each group, that is,

$$V = \left\{ x \in \{0, 1\}^n \,\middle|\, \forall j \in \{0, \ldots, c + 1\} \colon \sum_{i=1}^{n'} x_{j \cdot n' + i} = 1 \right\}.$$

Every virtual item $x^{(i)}$ is drawn independently and uniformly from the set V. It can happen that there exists an original item that occurs in more than k virtual items. In this case, the sequence of virtual items is not valid because we have only k copies of each item. Then the ranking is replaced by an arbitrary ranking on $\mathcal{D}^n$. The following lemma shows that this failure event is unlikely to occur.

Lemma 4 *The probability that the sequence of virtual items is not valid because more than k copies of one original item are contained in the virtual items is at most $1/(nk)^5$.*

Proof For $i \in [n]$, let L_i denote the number of instances of item i that are contained in the virtual items. We can bound the probability that L_i exceeds k by

$$\mathbf{Pr}[L_i \geq k + 1] \leq \binom{\ell}{k + 1} \cdot \left(\frac{1}{n'}\right)^{k+1}$$

$$\leq \left(\frac{e \cdot \ell}{k + 1}\right)^{k+1} \cdot \left(\frac{c + 2}{n}\right)^{k+1} < \left(\frac{1}{2}\right)^k.$$

A union bound yields

$$\mathbf{Pr}[\exists i \in [n] : L_i \geq k + 1] \leq n \cdot \left(\frac{1}{2}\right)^k \leq n \cdot \left(\frac{1}{2}\right)^{(5(c+1)+1)\log n}$$

$$= \frac{1}{n^{5(c+1)}} \leq \frac{1}{(nk)^5}.$$

$\qquad\qquad\Box$

We prove Theorem 2 in two steps. First we prove the following lemma about how the profits of the virtual items in V are distributed, where the profit of a virtual item $x \in \{0, 1\}^n$ is defined as $p \cdot x$. Observe that scaling all profits by the same factor does not change the number of Pareto-optimal solutions. Hence, we can assume that the profits are chosen uniformly at random from the interval $[-u, u]$ for an arbitrary $u > 0$.

Lemma 5 *If the profits $p_1, \ldots, p_n$ of the original items are chosen independently uniformly at random from the interval $[-n^{c+1}, n^{c+1}]$, then there exist constants $\gamma > 0$ and $p > 0$ depending only on c such that with probability at least p, for each $j \in \{0, \ldots, n^{c+1} - 1\}$, the set V contains at least n/γ virtual items whose profits lie in the interval $(j, j + 1)$.*

Let us remark that we scaled the profits of the original items to the interval $[-n^{c+1}, n^{c+1}]$ in the lemma above only to keep the notation less cumbersome. The benefit of scaling is that we can consider intervals $(j, j + 1)$ for integer values of j. Throughout the entire remainder of this section we will assume that the profits of the original items are chosen uniformly from $[-n^{c+1}, n^{c+1}]$ without mentioning this explicitly anymore. This applies in particular to the proof of Theorem 2 below. This assumption is without loss of generality because scaling all profits does not change the set of Pareto-optimal solutions.

Furthermore, we adapt the lower bound of $\Omega(n^2)$ in [5] for the binary case from uniformly random profits to profits that are chosen only "nearly" uniformly at random. To make this more precise, consider a knapsack instance with n items in which the ith item has weight 2^i and the profits of the items are chosen independently according to a probability distribution $F \colon \mathbb{R} \to \mathbb{R}_{\geq 0}$. Assume that F consists of two components, that is, there exists a constant $\delta > 0$ such that $F = \delta \cdot U + (1 - \delta) \cdot G$ for two probability distributions U and G. Furthermore, assume that U has the property that for each $j \in \{0, \ldots, T - 1\}$ it holds $\mathbf{Pr}[X \in (j, j + 1)] = 1/T$ for a random variable X distributed according to U and some $T \geq n$.

Lemma 6 *The expected number of Pareto-optimal solutions in the aforementioned scenario is at least $\delta^2 n^2 / 128$.*

Together Lemmas 4, 5, and 6 and the upper bound on the expected number of Pareto-optimal solutions presented in Theorem 1 imply Theorem 2.

Proof of Theorem 2 Assume that the ranking on the set of solutions is determined as described above, that is, the ranking is induced by ℓ randomly chosen virtual items from V. Let $\mathcal{F}_1$ denote the event that there exists some $j \in \{0, \ldots, n^{c+1} - 1\}$ for which less than n/γ elements in V have a profit in $(j, j + 1)$. Due to Lemma 5, the probability of the event $\mathcal{F}_1$ is at most $1 - p$. Intuitively, the failure event $\mathcal{F}_1$ occurs if the profits $p_1, \ldots, p_n$ are chosen such that the profit distribution of the virtual items is not uniform enough.

We first analyze the number q' of Pareto-optimal solutions in a different random experiment. In this random experiment, we do not care if the sequence of virtual items is valid, that is, we assume $\mathcal{D} = \{0, \ldots, \ell\}$, for which the sequence is always valid. We later use q' as an auxiliary random variable to analyze the number of Pareto-optimal solutions for the setting $\mathcal{D} = \{0, \ldots, k\}$, which we actually care about.

Claim It holds $\mathbf{E}\left[q' \mid \neg\mathcal{F}_1\right] \geq \kappa' n^2 k^2$ for $\kappa' = \delta^2/(512e^2(c+2)^2)$.

Proof Let X denote the random variable that describes the profit of a uniform random virtual item chosen from V. We argue in the following that under the assumption that $\mathcal{F}_1$ does not occur, we can write the distribution of X in a form such that Lemma 6 is applicable.

Under the assumption that $\mathcal{F}_1$ does not occur, for every $j \in \{0, \ldots, n^{c+1}-1\}$ at least n/γ virtual items have a profit in the interval $(j, j+1)$. For each interval $(j, j+1)$ we choose exactly n/γ virtual items with a profit in that interval arbitrarily. We call these virtual items good and all other virtual items bad. Altogether there are n^{c+2}/γ good items. Since there are $(n')^{c+2}$ virtual items in V in total, the probability that one of the good items is chosen is $\delta = n^{c+2}/(\gamma n'^{c+2}) = (c+2)^{c+2}/\gamma$. Under the condition that a good item is chosen, which happens with probability δ, the probability that X takes a value in the interval $(j, j+1)$ is exactly $1/n^{c+1}$ for every $j \in \{0, \ldots, n^{c+1}-1\}$. This corresponds to the distribution U in the setting of Lemma 6.

The number of virtual items in the sequence is $\ell = nk/(2e(c+2))$. These virtual items are chosen independently and if we assign the ith virtual item $x^{(i)}$ a weight of 2^i we can apply Lemma 6, yielding that the expected number of Pareto-optimal solutions of a knapsack instance with the virtual items is at least

$$\delta^2 \ell^2/128 = \kappa' n^2 k^2 \quad \text{for} \quad \kappa' = \delta^2/(512e^2(c+2)^2).$$

Altogether, we have shown $\mathbf{E}\left[q' \mid \neg\mathcal{F}_1\right] \geq \kappa' n^2 k^2$. $\square$

Observe that it can happen that there exist different subsets of the virtual items that represent the same original solution, i.e., there exist $I, J \subseteq \{1, \ldots, \ell\}$ with $I \neq J$ such that $\sum_{i \in I} x^{(i)} = \sum_{j \in J} x^{(j)}$. Due to the definition of $\mathcal{S}_i^*$ as $\mathcal{S}_i \setminus \mathcal{S}_{i-1}$, only the solution with highest ranking among these is considered in our construction. Since all solutions of the knapsack instance that represent the same original solution have the same profit only the one with the highest ranking can be a Pareto-optimal solution. Hence, leaving out the other solutions in our construction does not affect the number of Pareto-optimal solutions.

Now we take into account that the sequence of virtual items might not be a valid sequence for $\mathcal{D} = \{0, \ldots, k\}$ because more than k copies of one original item are contained in the virtual items. Let $\mathcal{F}_2$ denote the event that the sequence of virtual items is not allowed because it contains more than k instances of one item. Due to Lemma 4, we know that $\mathbf{Pr}[\mathcal{F}_2] \leq 1/(nk)^5$. Remember that if this failure event occurs, the ranking is set to an arbitrary ranking on $\mathcal{D}^n$. Let q denote the number of Pareto-optimal solutions. By definition of q' and the failure event $\mathcal{F}_2$, we know that $\mathbf{E}[q \mid \neg\mathcal{F}_2] = \mathbf{E}\left[q' \mid \neg\mathcal{F}_2\right]$. Furthermore, since $\mathcal{F}_2$ does not affect the choice of the profits, we can use Theorem 1 to bound $\mathbf{E}\left[q' \mid \mathcal{F}_2\right]$, but we have to take into account that in the modified random experiment for which q' is defined we have $\mathcal{D} = \{0, \ldots, \ell\}$. Hence, we obtain $\mathbf{E}\left[q' \mid \mathcal{F}_2\right] \leq \kappa'' n^4 k^2$ for a sufficiently large constant κ''.

Putting these results together yields

$$\mathbf{E}[q] \geq \mathbf{Pr}[\neg\mathcal{F}_2] \cdot \mathbf{E}[q \mid \neg\mathcal{F}_2]$$

$$= \mathbf{Pr}[\neg\mathcal{F}_2] \cdot \mathbf{E}[q' \mid \neg\mathcal{F}_2]$$

$$= \mathbf{E}[q'] - \mathbf{Pr}[\mathcal{F}_2] \cdot \mathbf{E}[q' \mid \mathcal{F}_2]$$

$$\geq \mathbf{Pr}[\neg\mathcal{F}_1] \cdot \mathbf{E}[q' \mid \neg\mathcal{F}_1] - \mathbf{Pr}[\mathcal{F}_2] \cdot \mathbf{E}[q' \mid \mathcal{F}_2]$$

$$\geq p \cdot \kappa' n^2 k^2 - \frac{\kappa'' n^4 k^2}{(nk)^5}$$

$$\geq \kappa n^2 k^2$$

for a sufficiently large constant κ. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

3.2.1 Proof of Lemma 5

In order to prove Lemma 5, we analyze an auxiliary random experiment first. A well-studied random process is the experiment of placing n balls uniformly and independently at random into m bins. In this random allocation process, the expected load of each bin is n/m and one can use Chernoff bounds to show that in the case $n \geq m$ it is unlikely that there exists a bin whose load deviates by more than a logarithmic factor from its expectation. In this section, we consider a random experiment in which the locations of the balls are chosen as linear combinations of independent random variables. Since the same random variables appear in linear combinations for different balls, the locations of the balls are dependent in a special way.

Let $c \in \mathbb{N}$ with $c \geq 2$ be an arbitrary constant and assume that we are given n independent random variables that are chosen uniformly at random from the interval $[-n^{c+1}, n^{c+1}]$. We assume that n is a multiple of $c + 2$, and we partition the set of random variables into $c + 2$ sets with $n' = n/(c + 2)$ random variables each. For $i \in \{1, \ldots, c + 2\}$ and $j \in \{1, \ldots, n'\}$, let p_j^i denote the j-th random variable in the ith group.

For every $\ell \in [c + 2]$, we consider a random experiment in which the set of balls is $[n']^{\ell}$ and the bins are the intervals $(-\ell n^{c+1}, -\ell n^{c+1} + 1), \ldots, (\ell n^{c+1} - 1, \ell n^{c+1})$. In the following, bin j denotes the interval $(j, j + 1)$. Hence, the number of balls is $(n')^{\ell}$ and the number of bins is $2\ell n^{c+1}$. Instead of placing these balls independently into the bins, the location of a ball $a \in [n']^{\ell}$ is chosen to be $p_{a_1}^1 + \cdots + p_{a_{\ell}}^{\ell}$, that is, it is placed in bin $\lfloor p_{a_1}^1 + \cdots + p_{a_{\ell}}^{\ell} \rfloor$. We will refer to this random process as round ℓ in the following. We show that despite these dependencies, the allocation process generates a more or less balanced allocation with constant probability. We use the following weighted Chernoff bound whose proof can be found in Appendix A.

Lemma 7 *Let* $X_1, \ldots, X_n$ *be independent discrete random variables with values in* $[0, z]$ *for some* $z > 0$. *Let* $X = \sum_{i=1}^n X_i$ *and* $\mu = \mathbf{E}[X]$. *Then for every* $x > 0$,

$$\mathbf{Pr}[X \geq x] < \left(\frac{e \cdot \mu}{x}\right)^{x/z} \quad \text{and} \quad \mathbf{Pr}[X \leq x] < \left(\frac{e^{1-\mu/x} \cdot \mu}{x}\right)^{x/z}.$$

We will first study how the random experiments for different values of ℓ are related. In round 1, there are n' balls that are placed at the positions $p_1^1, \ldots, p_{n'}^1$. These positions are chosen independently and uniformly at random from the interval $[-n^{c+1}, n^{c+1}]$. The bins correspond to the intervals $(j, j+1)$ for $j \in \{-n^{c+1}, n^{c+1} - 1\}$. Hence, each ball is placed in each bin with probability $1/(2n^{c+1})$ and the process in round 1 corresponds to the well-studied setting that n' balls are independently and uniformly at random allocated to $2n^{c+1}$ bins. We define $\mathcal{F}_1$ to be the event that there exists a bin that contains more than one ball after the first round. Since $c \geq 2$, a union bound over the $\binom{n'}{2} \leq (n')^2$ pairs of balls implies

$$\mathbf{Pr}[\mathcal{F}_1] \leq \frac{(n')^2}{2n^{c+1}} = o(1)$$

because the probability that two specific balls are assigned to the same bin is $1/(2n^{c+1})$.

Let $\ell \geq 2$. We can describe round ℓ as follows: Replace each ball from round $\ell - 1$ by n' identical copies. Then these n' copies are moved, where the location of the j-th copy is obtained by adding p_j^ℓ to the current location. Let X_j^ℓ denote the number of balls in bin j after the ℓth round.

For $\ell \in \{2, \ldots, c\}$, let $\mathcal{F}_\ell$ denote the event that after round ℓ there exists a bin that contains more than $(2c + 4)^{\ell-1}$ balls. Furthermore, let $\mathcal{F}_{c+1}$ denote the event that after round $c + 1$ there exists a bin that contains more than $\ln n$ balls.

Lemma 8 *After round $c+1$, the average number of balls per bin is a constant depending on c, and with probability $1 - o(1)$, the maximal number of balls in any bin is bounded from above by $\ln n$.*

Proof The number of balls in round $c + 1$ is $(n')^{c+1}$ and the number of bins is $2(c + 1)n^{c+1}$. Hence, the average number of balls per bin is

$$\frac{(n')^{c+1}}{2(c+1)n^{c+1}} = \frac{n^{c+1}}{2(c+1)n^{c+1}(c+2)^{c+1}} = \frac{1}{2(c+1)(c+2)^{c+1}} = \Theta(1).$$

Let $\ell \in \{2, \ldots, c\}$. Assume that the random variables in the first $\ell - 1$ groups are already fixed in such a way that the event $\mathcal{F}_{\ell-1}$ does not occur. Under this assumption, also the variables $X_j^{\ell-1}$ are fixed and have values of at most $(2c + 4)^{\ell-2}$. Consider a bin j after round $\ell - 1$ and assume that to all elements in that bin the d-th element of the ℓth group is added. The locations of the balls obtained this way are in the interval $(j + \lfloor p_d^\ell \rfloor, j + \lfloor p_d^\ell \rfloor + 2)$, that is, they lie either in bin $j + \lfloor p_d^\ell \rfloor$ or $j + \lfloor p_d^\ell \rfloor + 1$. Hence, we can bound X_j^ℓ by

$$X_j^\ell \leq \sum_{d=1}^{n'} Y_{j, p_d^\ell}^{\ell-1} \text{ with } Y_{j, p_d^\ell}^{\ell-1} := X_{j-\lfloor p_d^\ell \rfloor}^{\ell-1} + X_{j-\lfloor p_d^\ell \rfloor - 1}^{\ell-1}.$$

Hence, when the random variables in the first $\ell - 1$ groups are fixed such that $\mathcal{F}_{\ell-1}$ does not occur, then X_j^ℓ is bounded by the sum of independent discrete random variables

$Y_{j,p_d^\ell}^{\ell-1}$ that take only values from the set $\{0, \ldots, 2(2c+4)^{\ell-2}\}$. The expected value of X_j^ℓ is bounded from above by $(n')^\ell/(2n^{c+1}) < 1/n$. Altogether, this implies that we can use Lemma 7 to bound the probability that X_j^ℓ exceeds its expectation. We obtain

$$\mathbf{Pr}\left[X_j^\ell \geq (2c+4)^{\ell-1} \,\Big|\, \neg\mathcal{F}_{\ell-1}\right] \leq \left(\frac{e}{(2c+4)^{\ell-1}n}\right)^{c+2} < n^{-(c+2)}.$$

Applying a union bound over all $2\ell n^{c+1}$ bins j yields

$$\mathbf{Pr}\left[\mathcal{F}_\ell \mid \neg\mathcal{F}_{\ell-1}\right] \leq \mathbf{Pr}\left[\exists j : X_j^\ell \geq (2c+4)^{\ell-1} \,\Big|\, \neg\mathcal{F}_{\ell-1}\right] \leq (2\ell n^{c+1}) \cdot n^{-(c+2)} = o(1).$$

Now consider round $c+1$. The expected value of X_j^{c+1} is bounded from above by $(n')^{c+1}/(2n^{c+1}) < 1$ and the same arguments as for the previous rounds show that X_j^{c+1} can be bounded by the sum of independent random variables with values from the set $\{0, \ldots, 2(2c+4)^{c-1}\}$ when the random variables in the first c groups are fixed such that $\mathcal{F}_c$ does not occur. Hence, we can again apply Lemma 7 to obtain

$$\mathbf{Pr}\left[X_j^{c+1} \geq \ln n \,\Big|\, \neg\mathcal{F}_c\right] \leq \left(\frac{e}{\ln n}\right)^{\ln n/(2(2c+4)^{c-1})} \leq n^{-\ln\ln n/(2(2c+4)^{c-1})+1}.$$

Let $\mathcal{F}_{c+1}$ denote the event that after round $c+1$ there exists a bin that contains more than $\ln n$ balls. Applying a union bound over all $2(c+1)n^{c+1}$ bins yields

$$\mathbf{Pr}\left[\mathcal{F}_{c+1} \mid \neg\mathcal{F}_c\right] \leq (2(c+1)n^{c+1}) \cdot n^{-\ln\ln n/(2(2c+4)^{c-1})+1} = o(1).$$

Now we can bound the probability that $\mathcal{F}_{c+1}$ occurs as

$$\begin{aligned}
\mathbf{Pr}\left[\mathcal{F}_{c+1}\right] &\leq \mathbf{Pr}[\mathcal{F}_c] + \mathbf{Pr}\left[\mathcal{F}_{c+1} \mid \neg\mathcal{F}_c\right] = \mathbf{Pr}[\mathcal{F}_c] + o(1) \\
&\leq \mathbf{Pr}\left[\mathcal{F}_{c-1}\right] + \mathbf{Pr}\left[\mathcal{F}_c \mid \neg\mathcal{F}_{c-1}\right] + o(1) = \mathbf{Pr}\left[\mathcal{F}_{c-1}\right] + o(1) \\
&\leq \mathbf{Pr}\left[\mathcal{F}_{c-2}\right] + \mathbf{Pr}\left[\mathcal{F}_{c-1} \mid \neg\mathcal{F}_{c-2}\right] + o(1) = \mathbf{Pr}\left[\mathcal{F}_{c-2}\right] + o(1) \\
&\leq \ldots = \mathbf{Pr}[\mathcal{F}_1] + o(1) = o(1).
\end{aligned}$$

$\square$

Based on Lemma 8, we prove the following lemma about the allocation after round $c+2$, which directly implies Lemma 5. To see this implication, observe that the auxiliary random experiment that we analyze in this section corresponds exactly to the setting of Lemma 5, where the virtual items correspond to the balls in round $c+2$ and the intervals $(j, j+1)$ in Lemma 5 correspond to the bins. The random variables p_j^i for $i \in \{1, \ldots, c+2\}$ and $j \in \{1, \ldots, n'\}$ with $n' = n/(c+2)$ correspond to the profits $p_1, \ldots, p_n$ of the original items.

Lemma 9 *For every constant $c \geq 2$, there exist constants $\gamma > 0$ and $p > 0$ such that with probability at least p the above described process yields after round $c+2$*

an allocation of the $(n')^{c+2}$ balls to the $2(c+2)n^{c+1}$ bins in which every bin $j \in \{0, \ldots, n^{c+1} - 1\}$ contains at least n/γ balls.

Proof In order to analyze the last round, we need besides $\neg \mathcal{F}_{c+1}$ one additional property that has to be satisfied after round $c + 1$. Let Y denote the number of balls after round $c+1$ that are assigned to bins j with $j \in \{0, \ldots, n^{c+1} - 1\}$. The probability that a fixed ball $a \in [n']^{c+1}$ is placed in one of these bins is at least $1/(2(c+1))$. Hence, the expected value of Y is at least $(n')^{c+1}/(2(c+1))$. Let $\overline{Y}$ denote the number of balls after round $c + 1$ that are not assigned to bins in $\{0, \ldots, n^{c+1} - 1\}$. The expected value of $\overline{Y}$ is at most $(n')^{c+1}(2c + 1)/(2c + 2)$. Applying Markov's inequality yields

$$\mathbf{Pr}\left[Y \le \frac{(n')^{c+1}}{4c + 4}\right] \le \mathbf{Pr}\left[\overline{Y} \ge \mathbf{E}\left[\overline{Y}\right] \frac{4c + 3}{4c + 2}\right] \le \frac{4c + 2}{4c + 3}.$$

Let $\mathcal{G}$ denote the failure event that Y is less than $(n')^{c+1}/(4c + 4)$. We have seen that $\neg \mathcal{G}$ occurs with constant probability.

Now we analyze round $c + 2$ and assume that the random variables in the first $c + 1$ groups are fixed in such a way that $\neg \mathcal{F}_{c+1} \cap \neg \mathcal{G}$ occurs.

Claim Consider a bin $j \in \{0, \ldots, n^{c+1} - 1\}$. Under the assumption $\neg \mathcal{G}$, the expected value of X_j^{c+2} is at least $n/(c + 2)^{c+5}$.

Proof Under the assumption $\neg \mathcal{G}$, there are together at least $(n')^{c+1}/(4c + 4)$ balls in the bins in $\{0, \ldots, n^{c+1} - 1\}$ after round $c + 1$. Each of these balls is at a location from the interval $[0, n^{c+1}]$. Fix an arbitrary such ball at location $x \in [0, n^{c+1}]$. From this ball, we obtain n' balls in round $c + 2$ by adding one of the numbers $p_1^{c+2}, \ldots, p_{n'}^{c+2}$ to its current location x. Adding a number from $(j - x, j + 1 - x)$ to x results in a ball with location in $(j, j + 1)$. Since $j \in \{0, \ldots, n^{c+1} - 1\}$ and $x \in [0, n^{c+1}]$, the interval $(j - x, j + 1 - x)$ lies inside the interval $[-n^{c+1}, n^{c+1}]$ from which the numbers $p_1^{c+2}, \ldots, p_{n'}^{c+2}$ are chosen uniformly at random. Hence, for every i, we have $\mathbf{Pr}\left[p_i^{c+2} \in (j - x, j + 1 - x)\right] = 1/(2n^{c+1})$.

Since there are at least $(n')^{c+1}/(4c + 4)$ balls in the bins in $\{0, \ldots, n^{c+1} - 1\}$ and each of them gives rise to n' balls in round $c + 2$, the expected value of X_j^{c+2} is at least

$$\frac{(n')^{c+2}}{((4c + 4)2n^{c+1})} \ge \frac{n}{(c + 2)^{c+5}}. \qquad \square$$

Remember that $\mathcal{F}_{c+1}$ denotes the event that after round $c + 1$ there exists a bin that contains more than $\ln n$ balls. Hence, under the assumption $\neg \mathcal{F}_{c+1}$, the random variable

$$Y_{j,p_d^{c+2}}^{c+1} := X_{j - \lfloor p_d^{c+2}\rfloor}^{c+1} + X_{j - \lfloor p_d^{c+2}\rfloor - 1}^{c+1},$$

which we already used in the proof of Lemma 8, takes only values in the interval $\{0, \ldots, 2 \ln n\}$ because each X_i^{c+1} is at most $\ln n$.

We apply Lemma 7 to bound the probability that X_j^{c+2} deviates from its mean:

$$\mathbf{Pr}\left[X_j^{c+2} \leq \frac{n}{2(c+2)^{c+5}} \,\middle|\, \neg\mathcal{F}_{c+1} \cap \neg\mathcal{G} \right] \leq \left(\frac{2}{e}\right)^{n/(4(c+2)^{c+5} \ln n)}.$$

Let $\mathcal{F}$ denote the event that there exists a bin $j \in \{0, \ldots, n^{c+1} - 1\}$ whose load is smaller than $n/(2(c+2)^{c+5})$. We can bound the probability of $\mathcal{F}$ by

$$\mathbf{Pr}\left[\mathcal{F} \mid \neg\mathcal{F}_{c+1} \cap \neg\mathcal{G} \right] \leq n^{c+1} \cdot \left(\frac{2}{e}\right)^{n/(4(c+2)^{c+5} \ln n)} = o(1).$$

Altogether this implies

$$\begin{aligned}
\mathbf{Pr}[\mathcal{F}] &\leq \mathbf{Pr}\left[\mathcal{F}_{c+1} \cup \mathcal{G}\right] + \mathbf{Pr}\left[\mathcal{F} \mid \neg\mathcal{F}_{c+1} \cap \neg\mathcal{G}\right] \\
&\leq \mathbf{Pr}[\mathcal{F}_c] + \mathbf{Pr}\left[\mathcal{F}_{c+1} \mid \neg\mathcal{F}_c\right] + \frac{4c+2}{4c+3} + o(1) \\
&\leq \mathbf{Pr}\left[\mathcal{F}_{c-1}\right] + \mathbf{Pr}\left[\mathcal{F}_c \mid \neg\mathcal{F}_{c-1}\right] + \frac{4c+2}{4c+3} + o(1) \\
&\leq \mathbf{Pr}\left[\mathcal{F}_{c-2}\right] + \mathbf{Pr}\left[\mathcal{F}_{c-1} \mid \neg\mathcal{F}_{c-2}\right] + \frac{4c+2}{4c+3} + o(1) \\
&\leq \mathbf{Pr}[\mathcal{F}_1] + \mathbf{Pr}\left[\mathcal{F}_2 \mid \neg\mathcal{F}_1\right] + \frac{4c+2}{4c+3} + o(1) \\
&\leq \frac{4c+2}{4c+3} + o(1),
\end{aligned}$$

which yields the lemma. □

3.2.2 Proof of Lemma 6

Beier and Vöcking [5] prove a lower bound of $\Omega(n^2)$ on the expected number of Pareto-optimal knapsack fillings for exponentially growing weights and profits that are chosen independently and uniformly at random from the interval $[0, 1]$. In this section, we adapt their proof for a random experiment in which the profits are chosen only "nearly" uniformly at random. Assume that we are given n items and that the ith item has weight $w_i = 2^i$. Furthermore, let $T \in \mathbb{N}$ be given and assume that $T \geq n$. In order to determine the profit p_i of the ith item, first one of the intervals $(0, 1), (1, 2), \ldots, (T - 1, T)$ is chosen uniformly at random. Then an adversary is allowed to choose the exact profit within the randomly chosen interval. We call an item whose profit is chosen this way a *nearly uniform item*. We prove that also in this scenario the expected number of Pareto-optimal solutions is lower bounded by $\Omega(n^2)$.

Lemma 10 *For instances consisting of n nearly uniform items, the expected number of Pareto-optimal solutions is bounded from below by $n^2/16$.*

Proof The proof follows along the lines of the proof of Theorem 3 for the binary case. Let $\mathcal{P}_j$ denote the set of Pareto-optimal solutions over the first j items, and let P_j denote the total profit of the first j items. For $j \in [n]$ and $\alpha \geq 0$, let X_α^j denote the number of Pareto-optimal solutions from $\mathcal{P}_j$ with profits in the interval $[P_j - \alpha, P_j)$. Observe that $p_j > \alpha$ implies $X_\alpha^j = X_\alpha^{j-1}$ and $p_j < \alpha$ implies $X_\alpha^j = X_{p_j}^{j-1} + X_{\alpha - p_j}^{j-1} + 1$. For integral $\alpha \in [T]$, the adversary cannot influence the event $p_j < \alpha$, as the interval from which he is allowed to pick values for p_j lies either completely left or completely right of α. Hence, for $\alpha \in [T]$ we can bound the expected value of X_α^j recursively as follows:

$$
\mathbf{E}\left[X_\alpha^j\right] \geq \mathbf{Pr}\left[p_j > \alpha\right] \cdot \mathbf{E}\left[X_\alpha^{j-1} \mid p_j > \alpha\right]
$$
$$
+ \mathbf{Pr}\left[p_j < \alpha\right] \cdot \left(\mathbf{E}\left[X_{p_j}^{j-1} \mid p_j < \alpha\right] + \mathbf{E}\left[X_{\alpha - p_j}^{j-1} \mid p_j < \alpha\right] + 1\right).
$$

As X_β^{j-1} is independent of p_j and X_β^{j-1} is monotone in β, we have

$$
\mathbf{E}\left[X_\alpha^j\right] \geq \mathbf{Pr}\left[p_j > \alpha\right] \cdot \mathbf{E}\left[X_\alpha^{j-1}\right]
$$
$$
+ \mathbf{Pr}\left[p_j < \alpha\right] \cdot \left(\mathbf{E}\left[X_{\lfloor p_j \rfloor}^{j-1} \mid p_j < \alpha\right] + \mathbf{E}\left[X_{\lfloor \alpha - p_j \rfloor}^{j-1} \mid p_j < \alpha\right] + 1\right).
\tag{5}
$$

In the following, we prove by induction on j that for every $\alpha \in [T]$,

$$
\mathbf{E}\left[X_\alpha^j\right] \geq \frac{\alpha \cdot j}{2T}.
$$

For $j = 1$ and $\alpha \in [T]$, we obtain

$$
\mathbf{E}\left[X_\alpha^1\right] = \mathbf{Pr}[p_1 < \alpha] = \frac{\alpha}{T} \geq \frac{\alpha}{2T}.
$$

Using the induction hypothesis and (5), we obtain, for $j \in [n] \setminus \{1\}$ and $\alpha \in [T]$, that $\mathbf{E}\left[X_\alpha^j\right]$ is lower bounded by

$$
\mathbf{Pr}\left[p_j > \alpha\right] \cdot \mathbf{E}\left[X_\alpha^{j-1}\right]
$$
$$
+ \mathbf{Pr}\left[p_j < \alpha\right] \left(\mathbf{E}\left[X_{\lfloor p_j \rfloor}^{j-1} \mid p_j < \alpha\right] + \mathbf{E}\left[X_{\lfloor \alpha - p_j \rfloor}^{j-1} \mid p_j < \alpha\right] + 1\right)
$$
$$
\geq \frac{T - \alpha}{T} \cdot \frac{\alpha(j-1)}{2T} + \sum_{i=0}^{\alpha-1} \mathbf{Pr}\left[p_j \in (i, i+1)\right] \left(\mathbf{E}\left[X_i^{j-1}\right] + \mathbf{E}\left[X_{\alpha-i-1}^{j-1}\right] + 1\right)
$$
$$
\geq \frac{T - \alpha}{T} \cdot \frac{\alpha(j-1)}{2T} + \sum_{i=1}^{\alpha} \frac{1}{T} \left(\frac{i(j-1)}{2T} + \frac{(\alpha - i - 1)(j-1)}{2T} + 1\right)
$$

$$
\begin{aligned}
&= \frac{T-\alpha}{T} \cdot \frac{\alpha(j-1)}{2T} + \frac{\alpha}{T}\left(\frac{(\alpha-1)(j-1)}{2T} + 1\right) \\
&= \frac{\alpha(j-1)}{2T} + \frac{\alpha}{T}\left(1 - \frac{j-1}{2T}\right) \\
&\geq \frac{\alpha(j-1)}{2T} + \frac{\alpha}{T} \cdot \frac{1}{2} \\
&= \frac{\alpha \cdot j}{2T}.
\end{aligned}
$$

This yields the following lower bound on the expected number of Pareto-optimal solutions:

$$
\mathbf{E}[q] \geq \sum_{j=1}^{n} \mathbf{E}\left[X^{j-1}_{\lfloor p_j \rfloor}\right] \geq \sum_{j=1}^{n} \frac{\mathbf{E}\left[\lfloor p_j \rfloor\right] \cdot (j-1)}{2T} \geq \frac{T-1}{4T} \cdot \sum_{j=1}^{n}(j-1) \geq \frac{n^2}{16}.
$$

□

We further generalize the scenario that we considered above and analyze the expected number of Pareto-optimal solutions for instances that do not only consist of nearly uniform items but also of some adversarial items. To be more precise, we assume that the profit of each item is chosen as follows: First a coin is tossed which comes up head with probability $\delta > 0$. If the coin comes up head, then the profit of the item is chosen as for nearly uniform items, that is, an interval is chosen uniformly at random and after that an adversary may choose an arbitrary profit in that interval. If the coin comes up tail, then an arbitrary non-integer profit can be chosen by an oblivious adversary who does not know the outcomes of the previous profits.

Proof of Lemma 6 First of all, we show that the presence of adversarial items does not affect the lower bound for the expected number of Pareto-optimal solutions. That is, we show that if there are $\hat{n}$ nearly uniform items and an arbitrary number of adversarial items, one can still apply Lemma 10 to obtain a lower bound of $\hat{n}^2/16$ on the expected number of Pareto-optimal solutions. For this, consider the situation that the first j items are nearly uniform items and that item $j+1$ is an adversarial item. Due to Lemma 10, we obtain that the expected value of X^j_α is bounded from below by $j \cdot \alpha/(2T)$ for every $\alpha \in [T]$. We show that the expected value of X^{j+1}_α is lower bounded by the same value. For this, consider the two alternatives that the adversary has. He can either choose $p_{j+1} > \alpha$ or $p_{j+1} < \alpha$. In the former case, we have $X^j_\alpha = X^{j+1}_\alpha$. In the latter case, we have

$$
\mathbf{E}\left[X^{j+1}_\alpha\right] \geq \mathbf{E}\left[X^j_{\lfloor p_{j+1} \rfloor}\right] + \mathbf{E}\left[X^j_{\lfloor \alpha - p_j \rfloor}\right] + 1 \geq \frac{(\alpha-1)j}{2T} + 1 \geq \frac{\alpha \cdot j}{2T}.
$$

Hence, the adversarial profit of item $j+1$ does not affect the lower bound for the expected number of Pareto-optimal solutions. One can apply this argument inductively to show the desired lower bound of $\hat{n}^2/16$.

In expectation the number $\hat{n}$ of nearly uniform items is δn and applying a Chernoff bound yields that with high probability $\hat{n} \geq \delta n/2$. For sufficiently large n, we can bound the probability that $\hat{n} < \delta n/2$ from above by $1/2$. Hence, with probability $1/2$ the expected number of Pareto-optimal solutions is at least $(\delta n/2)^2/16$, and hence, the expected number of Pareto-optimal solutions is bounded from below by $(\delta n)^2/128$. □

A Weighted Chernoff bound

In this section, we prove Lemma 7 (which corresponds to Exercise 4.19 from [16]). Typically Chernoff bounds are formulated for sums of independent Poisson trials. In this section, we derive a Chernoff bound for general discrete random variables. The proof is based very closely on the one for sums of Poisson trials in [16]. In fact, the only part which needs to be exchanged is an upper bound on the moment generating function.

For a random variable X, let $M_X(t) = \mathbf{E}\left[e^{tX}\right]$ denote its *moment generating function*. Assume that X is the sum of independent random variables $X_1, \ldots, X_n$, where each X_i is a discrete random variable taking only values in $[0, 1]$. Fix an index i and consider the random variable X_i. Let $p : W \to \mathbb{R}_{\geq 0}$ be its distribution, where $W \subseteq [0, 1]$ is a countable set.

We can write the moment generating function of X_i as follows:

$$
\begin{aligned}
M_{X_i}(t) &= \mathbf{E}\left[e^{tX_i}\right] \\
&= \sum_{w \in W} p(w) \cdot e^{tw} \\
&\leq \sum_{w \in W} p(w) \cdot (we^t + 1 - w),
\end{aligned}
\tag{6}
$$

where the last inequality follows from the convexity of the function $f(x) = e^{tx}$ since

$$
e^{tw} = f(w) = f(w \cdot 1 + (1 - w) \cdot 0) \leq w \cdot f(1) + (1 - w) \cdot f(0) = w \cdot e^t + 1 - w.
$$

Inequality ([6](#6)) yields

$$
\begin{aligned}
M_{X_i}(t) &\le 1 + \sum_{w \in W} p(w) \cdot w \cdot (e^t - 1) \\
&= 1 + \mathbf{E}[X_i] \cdot (e^t - 1) \\
&\le \exp\left(\mathbf{E}[X_i] \cdot (e^t - 1)\right),
\end{aligned}
$$

where in the last inequality we have used the fact that, for any $y \in \mathbb{R}$, $1 + y \le e^y$.

Since the random variables $X_1, \ldots, X_n$ are assumed to be independent, the moment generating function of X is simply the product of the moment generating functions of the X_i's. Hence, we obtain

$$
\begin{aligned}
M_X(t) &= \prod_{i=1}^{n} M_{X_i}(t) \\
&\le \prod_{i=1}^{n} \exp\left(\mathbf{E}[X_i] \cdot (e^t - 1)\right) \\
&\le \exp\left(\sum_{i=1}^{n} \mathbf{E}[X_i] \cdot (e^t - 1)\right) \\
&= \exp\left(\mathbf{E}[X] \cdot (e^t - 1)\right).
\end{aligned}
$$

Now, we are ready to prove the following Chernoff bound.

Theorem 4 *Let $X_1, \ldots, X_n$ be independent discrete random variables with values in $[0, 1]$. Let $X = \sum_{i=1}^{n} X_i$ and $\mu = \mathbf{E}[X]$. Then for every $x > 0$,*

$$
\mathbf{Pr}[X \ge x] < \left(\frac{e \cdot \mu}{x}\right)^x.
$$

Proof Applying Markov's inequality, for any $\delta > 0$ and $t > 0$ we have

$$
\begin{aligned}
\mathbf{Pr}[X \ge (1+\delta)\mu] &= \mathbf{Pr}\left[e^{tX} \ge e^{t(1+\delta)\mu}\right] \\
&\le \frac{\mathbf{E}\left[e^{tX}\right]}{e^{t(1+\delta)\mu}} \\
&\le \frac{e^{(e^t - 1)\mu}}{e^{t(1+\delta)\mu}}.
\end{aligned}
$$

For any $\delta > 0$, we can set $t = \ln(1 + \delta) > 0$ to get

$$
\mathbf{Pr}[X \ge (1+\delta) \cdot \mu] \le \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^\mu.
$$

This yields the theorem since

$$\mathbf{Pr}[X \geq t] = \mathbf{Pr}\left[X \geq \left(1 + \left(\frac{x}{\mu} - 1\right)\right) \cdot \mu\right]$$
$$\leq \left(\frac{e^{x/\mu - 1}}{(x/\mu)^{x/\mu}}\right)^{\mu}$$
$$\leq \left(\frac{e \cdot \mu}{x}\right)^{x}.$$

$\square$

By an appropriate scaling, we obtain the following variant of Theorem 4.

Corollary 4 *Let $X_1, \ldots, X_n$ be independent discrete random variables with values in $[0, z]$ for some $z > 0$. Let $X = \sum_{i=1}^{n} X_i$ and $\mu = \mathbf{E}[X]$. Then for every $x > 0$,*

$$\mathbf{Pr}[X \geq x] < \left(\frac{e \cdot \mu}{x}\right)^{x/z}.$$

By similar calculations, one obtains the following corollary.

Corollary 5 *Let $X_1, \ldots, X_n$ be independent discrete random variables with values in $[0, z]$ for some $z > 0$. Let $X = \sum_{i=1}^{n} X_i$ and $\mu = \mathbf{E}[X]$. Then for every $x > 0$,*

$$\mathbf{Pr}[X \leq x] < \left(\frac{e^{1 - \mu/x} \cdot \mu}{x}\right)^{x/z}.$$

References

1. Beier, R. : Probabilistic analysis of discrete optimization problems. PhD thesis, Universität des Saarlandes (2004)
2. Brunsch, Tobias, Goyal, Navin, Rademacher, Luis, Röglin, Heiko: Lower bounds for the average and smoothed number of pareto-optima. Theor. Comput. **10**, 237–256 (2014)
3. Brunsch, Tobias, Röglin, Heiko: Improved smoothed analysis of multiobjective optimization. J. ACM **62**(1), 4:1-4:58 (2015)
4. Beier, René, Röglin, Heiko, Vöcking, Berthold: 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), pp. 53–67 (2007)
5. Beier, René, Vöcking, Berthold: Random knapsack in expected polynomial time. J. Comput. Syst. Sci. **69**(3), 306–329 (2004)
6. Corley, H William, Moon, IDouglas: Shortest paths in networks with vector weights. J. Optimiz. Theor. Appl. **46**(1), 79–86 (1985)
7. Diakonikolas, Ilias, Yannakakis, Mihalis: 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), pp. 74–88 (2007)
8. Ehrgott, Matthias: Gandibleux, Xavier: Multiobjective combinatorial optimization. In: Ehrgott, Matthias, Gandibleux, Xavier (eds.) Multiple criteria optimization-state of the art annotated bibliographic surveys, pp. 369–444. Kluwer Academic Publishers (2002)
9. Ehrgott, Matthias: Integer solutions of multicriteria network flow problems. Investig. Oper. **19**, 229–243 (1999)

10. Kellerer, Hans, Pferschy, Ulrich, Pisinger, David: Knapsack Problems. Springer (2004)
11. Klamroth, Kathrin, Wiecek, Margaret M.: Dynamic programming approaches to the multiple criteria knapsack problem. Naval Res. Logist. **47**(1), 57–76 (2000)
12. Mustafa, Adli, Goh, Mark: Finding integer efficient solutions for bicriteria and tricriteria network flow problems using dinas. Comput. Oper. Res. **25**(2), 139–157 (1998)
13. Müller-Hannemann, Matthias, Weihe, Karsten: Pareto shortest paths is often feasible in practice. In: Proceedings of the 5th international workshop on algorithm engineering (WAE), pages 185–198 (2001)
14. Moitra, Ankur, O'Donnell, Ryan: Pareto optimal solutions for smoothed analysts. SIAM J. Comput. **41**(5), 1266–1284 (2012)
15. Manthey, Bodo, Röglin, Heiko: Smoothed analysis: analysis of algorithms beyond worst case. Inf. Technol. **53**(6), 280–286 (2011)
16. Mitzenmacher, Michael, Upfal, Eli: Probability and Computing. Cambridge University Press (2005)
17. Nemhauser, George L., Ullmann, Zev: Discrete dynamic programming and capital allocation. Manag. Sci. **15**(9), 494–505 (1969)
18. Papadimitriou, Christos H., Yannakakis, Mihalis: 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), pp. 86–92 (2000)
19. Röglin, Heiko: Rösner, Clemens: the smoothed number of Pareto-optimal solutions in non-integer bicriteria optimization. In: Theory and applications of models of computation. volume 10185 of Lecture Notes in Computer Science, pp. 543–555. Springer, Cham (2017)
20. Röglin, Heiko, Teng, Shang-Hua: Smoothed analysis of multiobjective optimization. In: Proceedings of the 50th Annual IEEE symposium on foundations of computer science (FOCS), pp. 681–690 (2009)
21. Skriver, Anders JV., Andersen, Kim Allan: A label correcting approach for solving bicriterion shortest-path problems. Comput. Oper. Res. **27**(6), 507–524 (2000)
22. Spielman, Daniel A., Teng, Shang-Hua.: Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. J. ACM **51**(3), 385–463 (2004)
23. Spielman, Daniel A., Teng, Shang-Hua.: Smoothed analysis: an attempt to explain the behavior of algorithms in practice. Commun. ACM **52**(10), 76–84 (2009)
24. Vassilvitskii, Sergei, Yannakakis, Mihalis: Efficiently computing succinct trade-off curves. Theor. Comput. Sci. **348**(2), 334–356 (2005)