

Improved Smoothed Analysis of Multiobjective Optimization*

Tobias Brunsch Heiko Röglin

Department of Computer Science
University of Bonn, Germany
{brunsch,roeglin}@cs.uni-bonn.de

Abstract

We present several new results about smoothed analysis of multiobjective optimization problems. Motivated by the discrepancy between worst-case analysis and practical experience, this line of research has gained a lot of attention in the last decade. We consider problems in which d linear and one arbitrary objective function are to be optimized over a set $\mathcal{S} \subseteq \{0, 1\}^n$ of feasible solutions. We improve the previously best known bound for the smoothed number of Pareto-optimal solutions to $O(n^{2d} \phi^d)$, where ϕ denotes the perturbation parameter. Additionally, we show that for any constant c the c^{th} moment of the smoothed number of Pareto-optimal solutions is bounded by $O((n^{2d} \phi^d)^c)$. This improves the previously best known bounds significantly.

Furthermore, we address the criticism that the perturbations in smoothed analysis destroy the zero-structure of problems by showing that the smoothed number of Pareto-optimal solutions remains polynomially bounded even for zero-preserving perturbations. This broadens the class of problems captured by smoothed analysis and it has consequences for non-linear objective functions. One corollary of our result is that the smoothed number of Pareto-optimal solutions is polynomially bounded for polynomial objective functions. Our results also extend to integer optimization problems.

1 Introduction

In most real-life decision-making problems there is more than one objective to be optimized. For example, when booking a train ticket, one wishes to minimize the travel time, the fare, and the number of train changes. As different objectives are often conflicting, usually no solution is simultaneously optimal in all criteria and one has to make a trade-off between different objectives. The most common way to filter out unreasonable trade-offs and to reduce the number of solutions the decision maker has to choose from is to determine the set of *Pareto-optimal solutions*, where a solution is called Pareto-optimal if no other solution is simultaneously better in all criteria.

Multiobjective optimization problems have been studied extensively in operations research and theoretical computer science (see, e.g., [10] for a comprehensive survey). In particular, many algorithms for generating the set of Pareto-optimal solutions for various optimization problems such as the (bounded) knapsack problem [17, 13], the multiobjective shortest path problem [8, 12, 20], and the multiobjective network flow problem [9, 16] have been proposed. Enumerating the set of Pareto-optimal solutions is not only used as a preprocessing step to eliminate unreasonable trade-offs, but often it is also used as an intermediate step in algorithms for solving optimization problems. For example, the Nemhauser–Ullmann algorithm [17] treats the single-criterion knapsack problem as a bicriteria optimization problem in which a solution with small weight and large profit

*The paper appeared in a preliminary version in the proceedings of STOC 2012 and will appear in JACM.

is sought, and it generates the set of Pareto-optimal solutions, ignoring the given capacity of the knapsack. After this set has been generated, the algorithm returns the solution with the highest profit among all Pareto-optimal solutions with weight not exceeding the knapsack capacity. This solution is optimal for the given instance of the knapsack problem.

Generating the set of Pareto-optimal solutions (a.k.a. the *Pareto set*) only makes sense if few solutions are Pareto-optimal. Otherwise, it is too costly and it does not provide enough guidance to the decision maker. While, in many applications, it has been observed that the Pareto set is indeed usually small (see, e.g., [15] for an experimental study of the multiobjective shortest path problem), one can, for almost every problem with more than one objective function, find instances with an exponential number of Pareto-optimal solutions (see, e.g., [10]).

Motivated by the discrepancy between worst-case analysis and practical observations, *smoothed analysis* of multiobjective optimization problems has gained a lot of attention in the last decade. Smoothed analysis is a framework for judging the performance of algorithms that has been proposed in 2001 by Spielman and Teng [21] in order to explain why the simplex algorithm is efficient in practice even though it has an exponential worst-case running time. In this framework, inputs are generated in two steps: first, an adversary chooses an arbitrary instance, and then this instance is slightly perturbed at random. The smoothed performance of an algorithm is defined to be the worst expected performance the adversary can achieve. This model can be viewed as a less pessimistic worst-case analysis, in which the randomness rules out pathological worst-case instances that are rarely observed in practice but dominate the worst-case analysis. If the smoothed running time of an algorithm is low and inputs are subject to a small amount of random noise then it is unlikely to encounter an instance on which the algorithm performs poorly. In practice, random noise can stem from measurement errors, numerical imprecision or rounding errors. It can also model arbitrary influences, which we cannot quantify exactly, but for which there is also no reason to believe that they are adversarial.

After its invention in 2001, smoothed analysis has been successfully applied in a variety of contexts, e.g., to explain the practical success of local search methods, heuristics for the knapsack problem, online algorithms, and clustering. A recent survey by Spielman and Teng [22] summarizes some of these results. One of the areas in which smoothed analysis has been applied extensively is multiobjective optimization. In 2003 Beier and Vöcking [3] initiated this line of research by showing that the smoothed number of Pareto-optimal solutions is polynomially bounded for all linear binary optimization problems with two objective functions. This was the first rigorous explanation why heuristics for generating the set of Pareto-optimal solutions are successful in practice despite their bad worst-case behavior. In the last years, Beier and Vöcking's original result has been improved and extended significantly in a series of papers. A discussion of this work follows in the next section after the formal description of the model.

1.1 Model and Previous Work

We consider a very general model of multiobjective optimization problems. An instance of such a problem consists of $d + 1$ objective functions V^1, \dots, V^{d+1} that are to be optimized over a set $\mathcal{S} \subseteq \{0, \dots, \mathcal{K}\}^n$ of feasible solutions for some integer \mathcal{K} . While the set \mathcal{S} and the last objective function $V^{d+1}: \mathcal{S} \rightarrow \mathbb{R}$ can be arbitrary, the first d objective functions have to be linear of the form $V^t(x) = V_1^t x_1 + \dots + V_n^t x_n$ for $x = (x_1, \dots, x_n) \in \mathcal{S}$ and $t \in \{1, \dots, d\}$. We assume without loss of generality that all objectives are to be minimized and we call a solution $x \in \mathcal{S}$ *Pareto-optimal* if there is no solution $y \in \mathcal{S}$ which is at least as good as x in all of the objectives and even better than x in at least one. We will introduce this notion formally in Section 2. The set of Pareto-optimal solutions is called the *Pareto set*. We are interested in the size of this set. As a convention, we count distinct Pareto-optimal solutions that coincide in all objective values only once. Since we compare solutions based on their objective values, there is no need to consider more than one solution with exactly the same values.

If one is allowed to choose the set \mathcal{S} , the objective function V^{d+1} , and the coefficients of the linear objective functions arbitrarily, then even for $d = 1$, one can construct instances with an exponential number of Pareto-optimal solutions. For this reason Beier and Vöcking introduced the model of ϕ -smooth instances [3], in which an adversary can choose the set \mathcal{S} and the objective function V^{d+1} arbitrarily while he can only specify a probability density function $f_i^t: [-1, 1] \rightarrow [0, \phi]$ for each coefficient V_i^t according to which it is chosen independently of the other coefficients. This model is more general than Spielman and Teng’s original two-step model in which the adversary first chooses coefficients which are afterwards subject to Gaussian perturbations. In ϕ -smooth instances the adversary can additionally determine the type of noise. He could, for example, specify for each coefficient an interval of length $1/\phi$ from which it is chosen uniformly at random. The parameter $\phi \geq 1/2$ can be seen as a measure for the power of the adversary: the larger ϕ the more precisely he can specify the coefficients of the linear objective functions. The aforementioned example of uniform distributions in intervals of length $1/\phi$ shows that for $\phi \rightarrow \infty$ smoothed analysis becomes a worst-case analysis.

The smoothed number of Pareto-optimal solutions depends on the number n of integer variables, the maximum integer \mathcal{K} , and the perturbation parameter ϕ . It is defined to be the largest expected number of Pareto-optimal solutions the adversary can achieve by any choice of $\mathcal{S} \subseteq \{0, \dots, \mathcal{K}\}^n$, $V^{d+1}: \mathcal{S} \rightarrow \mathbb{R}$, and the densities $f_i^t: [-1, 1] \rightarrow [0, \phi]$. In the following we assume that the adversary has made arbitrary fixed choices for these entities. Then we can associate with every matrix $V \in \mathbb{R}^{d \times n}$ the number $\text{PO}(V)$ of Pareto-optimal solutions in \mathcal{S} when the coefficients V_i^t of the d linear objective functions take the values given in V . Assuming that the adversary has made worst-case choices for \mathcal{S} , V^{d+1} , and the densities f_i^t , the smoothed number of Pareto-optimal solutions is the expected value $\mathbf{E}_V[\text{PO}(V)]$, where the coefficients in V are chosen according to the densities f_i^t . For $c \geq 1$, we call $\mathbf{E}_V[\text{PO}^c(V)]$ the c -th moment of the smoothed number of Pareto-optimal solutions. Here we assume that the adversary has made worst-case choices for \mathcal{S} , V^{d+1} , and the densities f_i^t that maximize $\mathbf{E}_V[\text{PO}^c(V)]$ (in general, these are different from the choices that maximize $\mathbf{E}_V[\text{PO}(V)]$).

Beier and Vöcking [3] showed that for the binary bicriteria case (i.e., $\mathcal{K} = d = 1$) the smoothed number of Pareto-optimal solutions is $O(n^4\phi)$ and $\Omega(n^2)$. The upper bound was later simplified and improved by Beier et al. [2] to $O(n^2\phi)$. In his PhD thesis [1], Beier conjectured that the smoothed number of Pareto-optimal solutions is polynomially bounded in n and ϕ for $\mathcal{K} = 1$ and every constant d . This conjecture was proven by Röglin and Teng [18], who showed that for binary solutions and for any fixed $d \geq 1$, the smoothed number of Pareto-optimal solutions is $O((n^2\phi)^{f(d)})$, where the function f is roughly $f(d) = 2^d d!$. They also proved that for any constant c the c -th moment of the smoothed number of Pareto-optimal solutions is bounded by $O((n^2\phi)^{c \cdot f(d)})$. Moitra and O’Donnell [14] improved the bound for the smoothed number of Pareto-optimal solutions significantly to $O(n^{2d}\phi^{d(d+1)/2})$. However, it remained unclear how to improve the bound for the moments by their methods. Recently a lower bound of $\Omega(n^{d-1.5}\phi^d)$ for the smoothed number of Pareto-optimal solutions was proven [6].

1.2 Our Results

In this article, we present several new results about smoothed analysis of multiobjective binary and integer optimization problems. Besides general ϕ -smooth instances, we additionally consider the special case of *quasiconcave density functions*. This means that we assume that every coefficient V_i^t is chosen independently according to its own density function $f_i^t: [-1, 1] \rightarrow [0, \phi]$ with the additional requirement that for every density f_i^t there is a value $x_i^t \in [-1, 1]$ such that f_i^t is non-decreasing in the interval $[-1, x_i^t]$ and non-increasing in the interval $[x_i^t, 1]$. We do not think that this is a severe restriction because all natural perturbation models, like Gaussian or uniform perturbations, use quasiconcave density functions. Furthermore, quasiconcave densities capture the essence of a perturbation: each coefficient V_i^t has an unperturbed value x_i^t and the probability

that the perturbed coefficient takes a value z becomes smaller with increasing distance $|z - x_i^t|$. We will call these instances *quasiconcave ϕ -smooth instances* in the following.

Beier and Vöcking originally only considered ϕ -smooth instances for binary bicriteria optimization problems (i.e., for the case $\mathcal{K} = d = 1$). The above described canonical generalization of this model to binary multiobjective optimization problems, on which Röglin and Teng's [18] and Moitra and O'Donnell's results [14] are based, appears to be very general and flexible at the first glance. However, one aspect limits its applicability severely and makes it impossible to formulate certain multiobjective linear optimization problems in this model. The weak point of the model is that it assumes that every binary variable x_i appears in every linear objective function as it is not possible to set some coefficients V_i^t deterministically to 0.

Already Spielman and Teng [21] and Beier and Vöcking [4] observed that the zeros often encode an essential part of the combinatorial structure of a problem and they suggested to analyze *zero-preserving perturbations* in which it is possible to either choose a density f_i^t according to which the coefficient V_i^t is chosen or to set it deterministically to 0. Zero-preserving perturbations have been studied in [19] and [4] for analyzing smoothed condition numbers of matrices and the smoothed complexity of binary optimization problems. For the smoothed number of Pareto-optimal solutions no upper bounds are known that are valid for zero-preserving perturbations (except trivial worst-case bounds), and in particular the bounds proven in [18] and [14] do not seem to generalize easily to zero-preserving perturbations. In this article, we develop new techniques for analyzing the smoothed number of Pareto-optimal solutions that can also be used for analyzing zero-preserving perturbations.

Theorem 1. *For any $d \geq 1$, the smoothed number of Pareto-optimal solutions is $\mathcal{K}^{(d+1)^5} \cdot O(n^{d^3+d^2+d}\phi^d)$ for quasiconcave ϕ -smooth instances with zero-preserving perturbations and $\mathcal{K}^{(d+1)^5} \cdot O((n\phi)^{d^3+d^2+d})$ for general ϕ -smooth instances with zero-preserving perturbations.*

Let us remark that the bounds stated in Theorem 1 hold for any $\mathcal{K} \geq 1$ and not only for sufficiently large values of \mathcal{K} . This is why the factor $\mathcal{K}^{(d+1)^5}$ is outside of the O -notation. The O -notation only refers to the parameters n and ϕ . For constant \mathcal{K} like in the binary case the factor $\mathcal{K}^{(d+1)^5}$ is a constant for fixed d . In Section 1.3 we will present some applications of zero-preserving perturbations. We will see that they allow us not only to extend the smoothed analysis to linear multiobjective optimization problems that are not captured by the previous model without zero-preserving perturbations, but that they also enable us to bound the smoothed number of Pareto-optimal solutions in problems with *non-linear objective functions*. In particular, the number of Pareto-optimal solutions for multivariate polynomial objective functions can be bounded by Theorem 1. We say that a ϕ -smooth instance has polynomial objective functions if every objective function V^t , $t \in \{1, \dots, d\}$, is the weighted sum of monomials, where the adversary can specify a ϕ -bounded density on $[-1, 1]$ for every weight according to which it is chosen. Denote the total number of monomials by m and let Δ denote the maximum degree of the monomials. Then the following corollary holds.

Corollary 2. *For any $d \geq 1$, the smoothed number of Pareto-optimal solutions is $\mathcal{K}^{(d+1)^5\Delta} \cdot O(m^{d^3+d^2+d}\phi^d)$ for quasiconcave ϕ -smooth instances with polynomial objective functions. For general ϕ -smooth instances with polynomial objective functions the smoothed number of Pareto-optimal solutions is $\mathcal{K}^{(d+1)^5\Delta} \cdot O((m\phi)^{d^3+d^2+d})$.*

In addition to zero-preserving perturbations we also study the standard model of ϕ -smooth instances. We present significantly improved bounds for the smoothed number of Pareto-optimal solutions and the moments, answering two questions posed by Moitra and O'Donnell [14].

Theorem 3. *For any $d \geq 1$, the smoothed number of Pareto-optimal solutions is $\mathcal{K}^{2(d+1)^2} \cdot O(n^{2d}\phi^d)$ for quasiconcave ϕ -smooth instances and $\mathcal{K}^{2(d+1)^2} \cdot O(n^{2d}\phi^{d(d+1)})$ for general ϕ -smooth instances.*

The bound of Theorem 3 for quasiconcave ϕ -smooth instances improves the previously best known bound of $O(n^{2d}\phi^{d(d+1)/2})$ in the binary case (which is, however, valid also for non-quasiconcave densities) and it answers a question posed by Moitra and O'Donnell whether it is possible to improve the factor of $\phi^{d(d+1)/2}$ in their bound [14]. Together with the recent lower bound of $\Omega(n^{d-1.5}\phi^d)$ [6], which is also valid for quasiconcave density functions, this shows that the exponents of both n and ϕ are linear in d .

Theorem 4. *For any $d \geq 1$ and any constant $c \in \mathbb{N}$, the c -th moment of the smoothed number of Pareto-optimal solutions is $\mathcal{K}^{(c+1)^2(d+1)^2} \cdot O((n^{2d}\phi^d)^c)$ for quasiconcave ϕ -smooth instances and $\mathcal{K}^{(c+1)^2(d+1)^2} \cdot O((n^{2d}\phi^{d(d+1)})^c)$ for general ϕ -smooth instances.*

This answers a question in [14] whether it is possible to improve the bounds for the moments in [18] and it yields better concentration bounds for the smoothed number of Pareto-optimal solutions. Our results also have immediate consequences for the expected running times of various algorithms because most heuristics for generating the Pareto set of some problem (including the ones mentioned at the beginning of the introduction) have a running time that depends linearly or quadratically on the size of the Pareto set.

The straightforward extension of the Nemhauser-Ullmann algorithm [17] to the multiobjective knapsack problem has, for example, a running time of $\Theta(\sum_{i=0}^{n-1} |P_i|^2)$ on instances with n items where P_i denotes the Pareto set of the instance that consists only of the first i items. (For $d = 1$ the running time can be made linear in $|P_i|$ if the sets P_i are stored in sorted order.) Other examples are the extensions of the Bellman-Ford algorithm and the Floyd-Warshall algorithm to multiobjective shortest path problems (see, e.g., [11]) whose running times depend linearly (for $d = 1$) or quadratically (for $d > 1$) on the number of Pareto-optimal solutions in certain subproblems. The improved bounds on the smoothed number of Pareto-optimal solutions and the second moment of this number yield improved bounds on the smoothed running times of these and various other algorithms.

Note that our analysis also covers the general case when the set \mathcal{S} is an arbitrary subset of $\{-\mathcal{K}, \dots, \mathcal{K}\}^n$. In this case, consider the shifted set $\mathcal{S}' = \{x + u : x \in \mathcal{S}\} \subseteq \{0, \dots, 2\mathcal{K}\}$ for $u = (\mathcal{K}, \dots, \mathcal{K})$ and the functions $W^1, \dots, W^{d+1}: \mathcal{S}' \rightarrow \mathbb{R}$, defined as $W^t x = V^t x$ for $t = 1, \dots, d$ and $W^{d+1} x = V^{d+1}(x - u)$. The Pareto set with respect to \mathcal{S} and $\{V^1, \dots, V^{d+1}\}$ and the Pareto set with respect to \mathcal{S}' and $\{W^1, \dots, W^{d+1}\}$ are identical except for a shift of $(V^1 u, \dots, V^d u, 0)$ in the image space. Hence, the sizes of both sets are equal. All aforementioned results can be applied for \mathcal{S}' and $\{W^1, \dots, W^{d+1}\}$, so they also hold for \mathcal{S} and $\{V^1, \dots, V^{d+1}\}$ if one replaces \mathcal{K} by $2\mathcal{K}$.

1.3 Applications of Zero-preserving Perturbations

Let us first of all remark that we can assume that the adversarial objective V^{d+1} is injective. If not, then let v_1, \dots, v_ℓ be the values taken by V^{d+1} and let $\Delta = \min_{i \neq j} |v_i - v_j|$. Now, consider an arbitrary injective function $\delta: \mathcal{S} \rightarrow [0, \Delta)$ and define the new adversarial objective as $W^{d+1} x = V^{d+1} x + \delta(x)$. Obviously, this function is injective and it preserves the order of the solutions in \mathcal{S} . This means that if $V^{d+1} x < V^{d+1} y$ for $x, y \in \mathcal{S}$, then also $W^{d+1} x < W^{d+1} y$. Let x be a Pareto optimum with respect to \mathcal{S} and $\{V^1, \dots, V^{d+1}\}$ and let x_2, \dots, x_m , $m \geq 1$, be all the other solutions for which $V^k x_i = V^k x$ for all $k \in \{1, \dots, d+1\}$. These are all Pareto optima but, due to our convention, we only count them once. Without loss of generality let x be the solution that minimizes W^{d+1} among these solutions. Then x is also Pareto-optimal with respect to \mathcal{S} and $\{V^1, \dots, V^d, W^{d+1}\}$.

Before we give some applications of zero-preserving perturbations let us remark that in the bicriteria case, which was studied in [3], zero-preserving perturbations are not more powerful than other perturbations because they can be simulated by the right choice of $\mathcal{S} \subseteq \{0, \dots, \mathcal{K}\}^n$ and the objective function $V^2: \mathcal{S} \rightarrow \mathbb{R}$.

Assume, for example, that the adversary has chosen \mathcal{S} and V^2 and has decided that the first coefficient V_1^1 of the first objective function should be deterministically set to 0. Also assume without loss of generality that V^2 is injective. We can partition the set \mathcal{S} into classes of solutions that agree in all components except for the first one. This means that two solutions $x \in \mathcal{S}$ and $y \in \mathcal{S}$ belong to the same class if $x_i = y_i$ for all $i \in \{2, \dots, n\}$. All solutions in the same class have the same value in the first objective V^1 as they differ only in the binary variable x_1 , whose coefficient has been set to 0. We construct a new set of solutions \mathcal{S}' that contains for every class only the solution with smallest value in V^2 . One can verify that the number of Pareto-optimal solutions is the same with respect to \mathcal{S} and with respect to \mathcal{S}' because all solutions in $\mathcal{S} \setminus \mathcal{S}'$ are dominated by solutions in \mathcal{S}' . Then we transform the set $\mathcal{S}' \subseteq \{0, \dots, \mathcal{K}\}^n$ into a set $\mathcal{S}'' \subseteq \{0, \dots, \mathcal{K}\}^{n-1}$ by dropping the first component of every solution. Furthermore, we define a function $W^2: \mathcal{S}'' \rightarrow \mathbb{R}$ that assigns to every solution $x \in \mathcal{S}''$ the same value that V^2 assigns to the corresponding solution in \mathcal{S}' . One can verify that the Pareto set with respect to \mathcal{S}' and V^2 is identical with the Pareto set with respect to \mathcal{S}'' and W^2 . The only difference is that in the latter problem we have eliminated the coefficient that is deterministically set to 0. Such an easy reduction of zero-preserving perturbations to other perturbations does not seem to be possible for $d \geq 2$ anymore.

Path Trading

Berger et al. [5] study a model for routing in networks. In their model there is a graph $G = (V, E)$ whose vertex set V is partitioned into mutually disjoint sets V_1, \dots, V_k . We can think of G as the Internet graph whose vertices are owned and controlled by k different autonomous systems (ASs). We denote by $E_i \subseteq E$ the set of edges inside V_i . The graph G is undirected, and each edge $e \in E$ has a length $\ell_e \in \mathbb{R}_{\geq 0}$. The traffic is modeled by a set of requests, where each request is characterized by its source node $s \in V$ and its target node $t \in V$. The Border Gateway Protocol (BGP) determines for each request (s, t) the order in which it has to be routed through the ASs. We say that a path P from s to t is valid if it connects s to t and visits the ASs in the order specified by the BGP protocol. This means that the first AS has to choose a path P_1 inside V_1 from s to some node in V_1 that is connected to some node $v_2 \in V_2$. Then the second AS has to choose a path P_2 inside V_2 from v_2 to some node in V_2 that is connected to some node $v_3 \in V_3$ and so on. For simplicity, the costs of routing a packet between two ASs are assumed to be 0, whereas AS i incurs costs of $\sum_{e \in P_i} \ell_e$ for routing the packet inside V_i along path P_i . In the common *hot-potato routing*, every AS is only interested in minimizing its own costs for each request. To model this, there are k objective functions that map each valid path P to a cost vector $(C_1(P), \dots, C_k(P))$, where

$$C_i(P) = \sum_{e \in P \cap E_i} \ell_e \quad \text{for } i \in \{1, \dots, k\}.$$

In [5] the problem of *path trading* is considered. If there is only one request, then no AS has an incentive to deviate from the hot-potato strategy. The problem becomes more interesting if there are multiple requests that have to be satisfied. Consider, for example, the three ASs depicted in Figure 1 and assume that there are three requests (s_1, t_1) , (s_2, t_2) , and (s_3, t_3) . Moreover, assume that the BGP specifies that all requests from $s \in V_i$ to $t \in V_j$ shall be routed directly from AS i to AS j . If all ASs follow the hot-potato strategy, then they decide for the routes (s_1, u_1, w_2, t_1) , (s_2, u_2, w_3, t_2) , and (s_3, u_3, w_1, t_3) . Each AS i incurs costs of 1 for the request (s_i, t_i) and costs of 9 for the request (s_j, t_j) for which $t_j \in V_i$.

Now assume that AS i routes request (s_i, t_i) from s_i to v_i . Then it incurs costs of 2 (instead of 1) for this route, which is worse than if it had chosen the hot-potato route. However, if all ASs agree on this new strategy, then each AS i only incurs costs of 2 (instead of 9) for the request (s_j, t_j) for which $t_j \in V_i$. Hence, the total costs of each AS for satisfying the three requests (s_i, t_i) is 4 instead of 10.

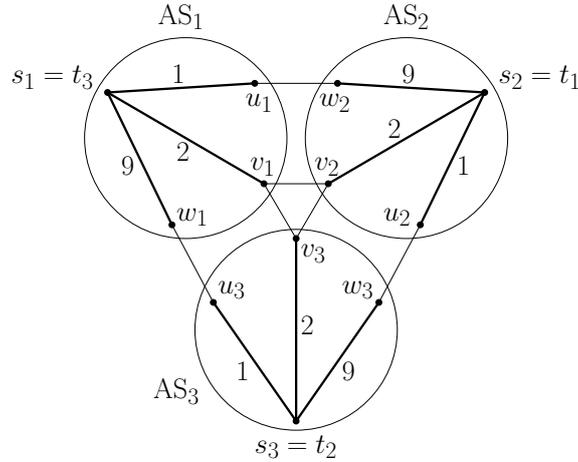


Figure 1: A network graph with three autonomous systems

The path trading problem asks whether there exist routes for given requests (s_i, t_i) such that the total costs of each involved AS is less than or equal to the total costs it would incur if all would follow the hot-potato strategy. Such routes are called *feasible path trades*.

Consider \mathcal{K} requests $(s_1, t_1), \dots, (s_{\mathcal{K}}, t_{\mathcal{K}})$ and s_i - t_i -paths $P_1, \dots, P_{\mathcal{K}}$ that comply with the BGP. For an edge $e \in E$ let $x_e \in \{0, \dots, \mathcal{K}\}$ be the number of paths $P_1, \dots, P_{\mathcal{K}}$ that contain e . We can encode the routes $P_1, \dots, P_{\mathcal{K}}$ by an integer vector $x \in \{0, \dots, \mathcal{K}\}^{|E|}$ consisting of the values x_e . Let \mathcal{S} denote the set of encodings of all valid routes $P_1, \dots, P_{\mathcal{K}}$. The question whether there is a feasible path trade for the requests (s_i, t_i) reduces to the question whether the vector x^* that encodes the hot-potato routes $P_1^*, \dots, P_{\mathcal{K}}^*$ is not Pareto-optimal with respect to \mathcal{S} and $\{C_1, \dots, C_k\}$, where the objectives $C_i: \mathcal{S} \rightarrow \mathbb{R}$,

$$C_i(x) = \sum_{e \in E_i} \ell_e x_e,$$

describe the total costs of AS i for the routes encoded by x . As the Pareto set can be exponentially large in the worst case, Berger et al. [5] proposed to study ϕ -smooth instances in which an adversary chooses the graph G and a density $f_e: [0, 1] \rightarrow [0, \phi]$ for every edge length ℓ_e according to which it is chosen. It seems as if we could easily apply the results in [18] and [14] to bound the smoothed number of Pareto-optimal paths because all objective functions C_i are linear in the binary variables x_e , $e \in E$. However, note that different objective functions contain different variables x_e because the coefficients of all x_e with $e \notin E_i$ are set to 0 in C_i . This is an important combinatorial property of the path trading problem that has to be obeyed. In the model in [18] and [14] it is not possible to set coefficients deterministically to 0. In their model, an AS would, with a probability of 1, incur positive costs for all edges and not only for its own edges that are used, which does not resemble the structure of the problem. Theorem 1, which allows zero-preserving perturbations, yields immediately the following result.

Corollary 5. *The smoothed number of Pareto-optimal valid paths is polynomially bounded in $|E|$, ϕ , and \mathcal{K} for any constant k .*

Non-linear Objective Functions

Even though we assumed above that the objective functions V^1, \dots, V^d are linear, we can also extend the smoothed analysis to non-linear objective functions. We consider first the bicriteria case $d = 1$. As above, we assume that the adversary has chosen an arbitrary set \mathcal{S} of feasible solutions

and an arbitrary injective objective function $V^2: \mathcal{S} \rightarrow \mathbb{R}$. In addition to that the adversary can choose m_1 arbitrary functions $I_i^1: \mathcal{S} \rightarrow \{0, \dots, \mathcal{K}\}$, $i \in \{1, \dots, m_1\}$. The objective function $V^1: \mathcal{S} \rightarrow \mathbb{R}$ is defined to be a weighted sum of the functions I_i^1 :

$$V^1(x) = \sum_{i=1}^{m_1} w_i^1 I_i^1(x),$$

where each weight w_i^1 is randomly chosen according to a density $f_i^1: [-1, 1] \rightarrow [0, \phi]$ given by the adversary. There is a wide variety of functions $V^1(x)$ that can be expressed in this way. We can, for example, express every polynomial if we let $I_1^1, \dots, I_{m_1}^1$ be its monomials. Note that the value \mathcal{K} then depends on the set \mathcal{S} and the maximum degree of the monomials.

We can linearize the problem by introducing a binary variable for every function I_i^1 . Using the function $\pi: \mathcal{S} \rightarrow \{0, \dots, \mathcal{K}\}^{m_1}$, defined by $\pi(x) = (I_1^1(x), \dots, I_{m_1}^1(x))$, the set of feasible solutions becomes $\mathcal{S}' = \{\pi(x) : x \in \mathcal{S}\} \subseteq \{0, \dots, \mathcal{K}\}^{m_1}$. For this set of feasible solutions we define $W^1: \mathcal{S}' \rightarrow \mathbb{R}$ and $W^2: \mathcal{S}' \rightarrow \mathbb{R}$ as follows:

$$W^1(y) = \sum_{i=1}^{m_1} w_i^1 y_i \quad \text{and} \quad W^2(y) = \min \{V^2(x) : x \in \mathcal{S} \text{ and } \pi(x) = y\}.$$

The problem defined by \mathcal{S} , V^1 , and V^2 and the problem defined by \mathcal{S}' , W^1 , and W^2 are equivalent and have the same number of Pareto-optimal solutions. The latter problem is linear and hence we can apply the result by Beier et al. [2], which yields that the smoothed number of Pareto-optimal solutions is bounded by $\text{poly}(\mathcal{K}) \cdot O(m_1^2 \phi)$. This shows in particular that the smoothed number of Pareto-optimal solutions is polynomially bounded in the number of monomials, the maximum integer in the monomials' ranges, and the density parameter for every polynomial objective function V^1 .

We can easily extend these considerations to multiobjective problems with $d \geq 2$. For these problems the adversary chooses an arbitrary set \mathcal{S} , numbers $m_1, \dots, m_d \in \mathbb{N}$, and an arbitrary injective objective function $V^{d+1}: \mathcal{S} \rightarrow \mathbb{R}$. In addition to that he chooses arbitrary functions $I_i^t: \mathcal{S} \rightarrow \{0, \dots, \mathcal{K}\}$ for $t \in \{1, \dots, d\}$ and $i \in \{1, \dots, m_t\}$. Every objective function $V^t: \mathcal{S} \rightarrow \mathbb{R}$ is a weighted sum

$$V^t(x) = \sum_{i=1}^{m_t} w_i^t I_i^t(x)$$

of the functions I_i^t , where each weight w_i^t is randomly chosen according to a density $f_i^t: [-1, 1] \rightarrow [0, \phi]$ chosen by the adversary. Similar to the bicriteria case, also this problem can be linearized. However, the previous results about the smoothed number of Pareto-optimal solutions can only be applied if every objective function V^t is composed of exactly the same functions I_i^t . Theorem 1 implies that the smoothed number of Pareto-optimal solutions is polynomially bounded in $\sum m_i$, \mathcal{K} , and ϕ , for any choice of the I_i^t .

Outline

After introducing some notation in the next section, we present an outline of our approach and our methods in Section 3. In our analysis we will frequently draw upon fundamental properties of Pareto-optimal solutions. These are stated and proven in Section 4. In Section 5 we prove Theorems 3 and 4. In Section 6 we consider zero-preserving perturbations and prove Theorem 1. We conclude the article with some open questions.

2 Notation

For the sake of simplicity we write $V^t x$ instead of $V^t(x)$, even for the adversarial objective V^{d+1} . With $V^{k_1 \dots k_t} x$ we refer to the vector $(V^{k_1} x, \dots, V^{k_t} x)$. In our analysis, we will shift the solutions

$x \in \mathcal{S}$ by a certain vector $u \in \{0, \dots, \mathcal{K}\}^n$ and consider the values $V^t \cdot (x - u)$. For the linear objectives we mean the value $V^t x - V^t u$, where $V^t u$ is well-defined even for a shift vector $u \in \{0, \dots, \mathcal{K}\}^n \setminus \mathcal{S}$. For the adversarial objective, however, we define $V^{d+1} \cdot (x - u) := V^{d+1} x$. It should not be confused with $V^{d+1} y$ for $y = x - u$. Note that for Pareto-optimality only the ordering of the solutions with respect to V^{d+1} and not the values $V^{d+1} x$ themselves are of interest. By the definition of $V^{d+1} \cdot (x - u)$, the ordering of the vectors $x - u$, $x \in \mathcal{S}$, equals the ordering of the vectors $x \in \mathcal{S}$ when considering V^{d+1} .

In the whole article let $\varepsilon > 0$ be an arbitrary real for which $1/\varepsilon$ is integral. Our analyses are valid for all such choices of ε , but to obtain our results we will consider the limit $\varepsilon \rightarrow 0$. Thus, think of ε as a very small real. Let $b = (b_1, \dots, b_d) \in \mathbb{R}^d$ be a vector such that b_k is an integral multiple of ε for all k . We will call the set $B = \{(y_1, \dots, y_d) \in \mathbb{R}^d : y_k \in (b_k, b_k + \varepsilon] \text{ for all } k\}$ an ε -box and b the corner of B . For a vector $x \in \{-\mathcal{K}, \dots, \mathcal{K}\}^n$ the expression $B_V(x)$ denotes the unique ε -box B for which $V^{1 \dots d} x \in B$. We call B the ε -box of x and say that x lies in B . With \mathbb{B}_ε we denote the set of all ε -boxes having corners b for which $b \in \{-n\mathcal{K}, -n\mathcal{K} + \varepsilon, \dots, n\mathcal{K} - 2\varepsilon, n\mathcal{K} - \varepsilon\}^d$. Hence, $|\mathbb{B}_\varepsilon| = (2n\mathcal{K}/\varepsilon)^d$. If all coefficients V_i^k of V are from $[-1, 1]$, which is true for all models considered in this article, and if for all $k = 1, \dots, d$ there is an index i such that $|V_i^k| < 1$, which holds with probability 1 in all of our models, then the ε -box of any vector $x \in \{-\mathcal{K}, \dots, \mathcal{K}\}^n$ belongs to \mathbb{B}_ε . Note that all vectors x constructed in this article are from $\{-\mathcal{K}, \dots, \mathcal{K}\}^n$. Hence, without any further explanation we will assume that $B_V(x) \in \mathbb{B}_\varepsilon$.

In this article we extensively use tuples instead of sets. The reason for this is that we are not only interested in certain components of a vector or matrix, but we also want to describe in which order they are considered. This will be clear after the introduction of the following notation. Let n, m be positive integers and let $a_1, \dots, a_n, b_1, \dots, b_m$ be arbitrary and not necessarily pairwise distinct reals. We define $[n] = (1, \dots, n)$, $[n]_0 = (0, 1, \dots, n)$, $|(a_1, \dots, a_n)| = n$ and $(a_1, \dots, a_n) \cup (b_1, \dots, b_m) = (a_1, \dots, a_n, b_1, \dots, b_m)$. By $(a_1, \dots, a_n) \setminus (b_1, \dots, b_m)$ and $(a_1, \dots, a_n) \cap (b_1, \dots, b_m)$ we denote the tuples we obtain by removing all occurrences of elements from (a_1, \dots, a_n) that do/do not belong to (b_1, \dots, b_m) . We write $(a_1, \dots, a_n) \subseteq (b_1, \dots, b_m)$ if $m \geq n$ and if (a_1, \dots, a_n) can be obtained from (b_1, \dots, b_m) by removing $m - n$ elements.

Let x be a vector and let A be a matrix. By $x|_{i_1 \dots i_n} = x|_{(i_1, \dots, i_n)}$ we denote the column vector $(x_{i_1}, \dots, x_{i_n})^T$, by $A|_{(i_1, \dots, i_n)}$ we denote the matrix consisting of the rows i_1, \dots, i_n of matrix A (in this order).

For an index set $I \subseteq [n]$ and a vector $y \in \{0, \dots, \mathcal{K}\}^n$ let $\mathcal{S}_I(y)$ denote the set of all solutions $z \in \mathcal{S}$ such that $z_i = y_i$ for all indices $i \in I$. For the sake of simplicity we also use the notation $\mathcal{S}_I(\hat{y})$ to describe the set $\{z \in \mathcal{S} : z_i = \hat{y}_i \text{ for all } i \in I\}$ for a vector $\hat{y} \in \{0, \dots, \mathcal{K}\}^{|I|}$ when the components of y are labeled by $y_{i_1}, \dots, y_{i_{|I|}}$ where $I = (i_1, \dots, i_{|I|})$.

With \mathbb{I}_n we refer to the $n \times n$ -identity matrix and with $\mathbb{O}_{m \times n}$ to the $m \times n$ -matrix whose entries are all 0. If the number of rows and columns are clear, then we drop the indices.

For a set $M \subseteq \mathbb{R}^n$ and a vector $y \in \mathbb{R}^n$ we define $M + y := \{x + y : x \in M\}$, the Minkowski sum of M and $\{y\}$.

Definition 6. Let $\mathcal{S} \subseteq \mathbb{R}^n$ be a set of solutions and let $f_1, \dots, f_d : \mathcal{S} \rightarrow \mathbb{R}$ be functions.

1. Let $x, y \in \mathbb{R}^n$ be vectors. We say that x *dominates* y (with respect to $\{f_1, \dots, f_d\}$), if $f_i(x) \leq f_i(y)$ for all $i \in [d]$ and $f_i(x) < f_i(y)$ for at least one $i \in [d]$. We say that x *dominates y strongly* (with respect to $\{f_1, \dots, f_d\}$), if $f_i(x) < f_i(y)$ for all $i \in [d]$.
2. Let $x \in \mathbb{R}^n$ be a vector. We call x *Pareto-optimal* or a *Pareto-optimum* (with respect to \mathcal{S} and $\{f_1, \dots, f_d\}$), if x is an element of \mathcal{S} and if no solution $y \in \mathcal{S}$ dominates x . We call x *weakly Pareto-optimal* or a *weak Pareto-optimum* (with respect to \mathcal{S} and $\{f_1, \dots, f_d\}$), if x is an element of \mathcal{S} and if no solution $y \in \mathcal{S}$ dominates x strongly.

We focus on Pareto-optimal solutions. The notions of strong dominance and weak Pareto-optimality are merely used for zero-preserving perturbations.

3 Outline of our Approach

To prove our results we adapt and improve methods from the previous analyses by Moitra and O'Donnell [14] and by Röglin and Teng [18] and combine them in a novel way. Since all coefficients of the linear objective functions lie in the interval $[-1, 1]$, for every solution $x \in \mathcal{S}$ the vector $V^{1 \dots d} x$ lies in the hypercube $[-n\mathcal{K}, n\mathcal{K}]^d$. The first step is to partition this hypercube into ε -boxes. If ε is very small (exponentially small in n), then it is unlikely that there are two different solutions $x \in \mathcal{S}$ and $y \in \mathcal{S}$ that lie in the same ε -box B unless x and y differ only in positions that are not perturbed in any of the objective functions, in which case we consider them as the same solution. In the remainder of this section we assume that no two solutions lie in the same ε -box. Then, in order to bound the number of Pareto-optimal solutions, it suffices to count the number of non-empty ε -boxes.

In order to prove Theorem 3 we show that for each fixed ε -box the probability that it contains a Pareto-optimal solution is bounded by $k \cdot \mathcal{K}^{2d^2+2d+1} n^d \phi^d \varepsilon^d$ for the constant $k = 2^{2d^2+3d+1} \cdot (d \cdot (d+1))^{d^2}$ that is hidden in the O -notation. This implies the theorem as the number of ε -boxes is $(2n\mathcal{K}/\varepsilon)^d$ and the exponent of \mathcal{K} is $2d^2 + 3d + 1 \leq 2(d+1)^2$. Fix an arbitrary ε -box B . In the following we will call a solution $x \in \mathcal{S}$ a *candidate* if there is a realization of V such that x is Pareto-optimal and lies in B . If there was only a single candidate $x \in \mathcal{S}$, then we could bound the probability that there is a Pareto-optimal solution in B by the probability that this particular solution x lies in B . This probability can easily be bounded from above by $\varepsilon^d \phi^d$ in the non-zero-preserving case. However, in principle, every solution $x \in \mathcal{S}$ can be a candidate and a union bound over all of them leads to a factor of $|\mathcal{S}|$ in the bound, which we have to avoid.

Following ideas of Moitra and O'Donnell, we divide the draw of the random matrix V into two steps. In the first step some information about V is revealed that suffices to limit the set of candidates to a single solution $x \in \mathcal{S}$. The exact position $V^{1 \dots d} x$ of this solution is determined in the second step. If the information that is revealed in these two steps is chosen carefully, then there is enough randomness left in the second step to bound the probability that x lies in the ε -box B . In Moitra and O'Donnell's analysis the coefficients in the matrix V are partitioned into two groups. In the first step the first group of coefficients is drawn, which suffices to determine the unique candidate x , and in the second step the remaining coefficients are drawn, which suffices to bound the probability that x lies in B . The second part consists essentially of $d(d+1)/2$ coefficients, which causes the factor of $\phi^{d(d+1)/2}$ in their bound.

We improve the analysis by a different choice of how to break the draw of V into two parts. As in the previous analysis, most coefficients are drawn in the first step. Only d^2 coefficients of V are drawn in the second step. However, these coefficients are not left completely random as in [14] because after the other coefficients have been drawn there can still be multiple candidates for Pareto-optimal solutions in B . Instead, the randomness is reduced further by drawing $d(d-1)$ linear combinations of these random variables in the first step. These linear combinations have the property that, after they have been drawn, there is a unique candidate x whose position can be described by d linear combinations that are linearly independent of the linear combinations already drawn in the first step. In [18] it was observed that linearly independent linear combinations of independent random variables behave in some respect similar to independent random variables. With this insight one can argue that in the second step there is still enough randomness to bound the probability that x lies in B . While the analysis in [18] yields only a bound proportional to $\phi^{d^2} \varepsilon^d$, we prove an improved result for quasiconcave densities that yields the desired bound proportional to $\phi^d \varepsilon^d$ (see Theorem 40).

In order to bound the c^{th} moment, we sum the probability that all ε -boxes B_1, \dots, B_c simultaneously contain a Pareto-optimal solution over all c -tuples (B_1, \dots, B_c) of ε -boxes. We bound this probability from above by $k \cdot \mathcal{K}^{c^2(d+1)^2+cd^2} n^{cd} \phi^{cd} \varepsilon^{cd}$ for the constant $k = 2^{c^2(d+1)^2+cd^2+cd} \cdot (cd(d+1))^{cd^2}$ that is hidden in the O -notation. Since there are $(2n\mathcal{K}/\varepsilon)^{cd}$ different c -tuples of ε -boxes and the exponent of \mathcal{K} is $c^2(d+1)^2 + cd^2 + cd \leq (c+1)^2(d+1)^2$, this implies the bound of

$\mathcal{K}^{(c+1)^2(d+1)^2} \cdot O((n^2\phi)^{cd})$ for the smoothed c^{th} moment of the number of Pareto-optimal solutions.

Let us fix a c -tuple (B_1, \dots, B_c) of ε -boxes. The approach to bound the probability that all of these ε -boxes contain simultaneously a Pareto-optimal solution is similar to the approach for the first moment. We divide the draw of V into two steps. In the first step enough information is revealed to identify for each of the ε -boxes B_i a unique candidate $x_i \in \mathcal{S}$ for a Pareto-optimal solution in B_i . If we do this carefully, then there is enough randomness left in the second step to bound the probability that $V^{1\dots d}x_i \in B_i$ for every $i \in [c]$. Again most coefficients are drawn in the first step and some linear combinations of the other cd^2 coefficients are also drawn in the first step. However, we cannot simply repeat the construction for the first moment independently c times because then there might be dependencies between the events $V^{1\dots d}x_i \in B_i$ for different i . In order to bound the probability that in the second step all x_i lie in their corresponding ε -boxes B_i , we need to ensure that the events $V^{1\dots d}x_i \in B_i$ are (almost) independent after the information from the first step has been revealed.

The general approach to handle zero-preserving perturbations is closely related to the approach for bounding the first moment for non-zero-preserving perturbations. However, additional complications have to be handled. The main problem is that we cannot easily guarantee anymore that the linear combinations in the second step are linearly independent of the linear combinations revealed in the first step. Essentially, the revealed linear combinations describe the positions of some solutions, which we will call *auxiliary solutions* in the following. For non-zero-preserving perturbations revealing this information is not critical as no solution has in any objective function exactly the same value as x . For zero-preserving solutions it can, however, happen that the auxiliary solutions take exactly the same value as x in one of the objective functions. Then there is not enough randomness left in the second step anymore to bound the probability that x lies in this objective in the ε -interval described by the ε -box B .

In the remainder of this section we will present some more details on our analysis. We first present a simplified argument to bound the smoothed number of Pareto-optimal solutions. Afterwards we will briefly discuss which changes to this argument are necessary to bound higher moments and to analyze zero-preserving perturbations.

Smoothed Number of Pareto-optimal Solutions As an important building block in the proof of Theorem 3 we use an insight from [14] about how to test whether a given ε -box contains a Pareto-optimal solution. Let us fix an ε -box $B = (b_1, b_1 + \varepsilon] \times \dots \times (b_d, b_d + \varepsilon]$ with corner $b = (b_1, \dots, b_d)$. The following algorithm takes as parameters the matrix V and the ε -box B and it returns a solution $x^{(0)}$.

Witness(V, B)

- 1: Set $\mathcal{R}_{d+1} = \mathcal{S}$.
- 2: **for** $t = d, d-1, \dots, 0$ **do**
- 3: Set $\mathcal{C}_t = \{z \in \mathcal{R}_{t+1} : V^{1\dots t}z \leq b|_{1\dots t}\}$.
- 4: Set $x^{(t)} = \arg \min \{V^{t+1}z : z \in \mathcal{C}_t\}$.
- 5: Set $\mathcal{R}_t = \{z \in \mathcal{R}_{t+1} : V^{t+1}z < V^{t+1}x^{(t)}\}$.
- 6: **end for**
- 7: **return** $x^{(0)}$

The actual **Witness** function that we use in the proof of Theorem 3 is more complex because it has to deal with some technicalities. In particular, the case that some set \mathcal{C}_t is empty, in which $x^{(t)}$ and \mathcal{R}_t would be undefined in the function above, has to be handled. For the purpose of illustration we ignore these technicalities here and assume that \mathcal{C}_t is never empty. The crucial observation that has been made by Moitra and O'Donnell is that if there is a Pareto-optimal solution $x \in \mathcal{S}$ that lies in B , then $x^{(0)} = x$ (assuming that no two solutions lie in the same ε -box). Hence, the solution $x^{(0)}$ returned by the **Witness** function is the only candidate for a Pareto-optimal solution in B . Our goal is to execute the **Witness** function and to obtain the solution $x^{(0)}$ without revealing the entire

matrix V . We will see that it is indeed possible to divide the draw of V into two steps such that in the first step enough information is revealed to execute the **Witness** function and such that in the second step there is still enough randomness left to bound the probability that $x^{(0)}$ lies in B .

We want to illustrate the case $d = 2$, in which there are one adversarial and two linear objective functions (even though the following reasoning is true for all $d \in \mathbb{N}$). For this, assume that B contains a single solution x which is Pareto-optimal and that x is very close to the corner b of B which can be assumed if B is very small. Then $V^t z \leq b_t$ is equivalent to $V^t z < V^t x$ for each $t \in [d]$.

Consider the situation depicted in Figure 2a. The first and the second objective value of each solution determine a point in the Euclidean plane. The additional value depicted next to this point represents the third objective value of each solution. Let us consider the situation before entering the loop. All points in Figure 2a are encircled meaning that \mathcal{R}_3 contains all solutions, i.e., $\mathcal{R}_3 = \mathcal{S}$. Now let us analyze the loop. The set \mathcal{C}_2 contains all solutions that have smaller first and second objective values than x (gray area in Figure 2b). Among these solutions we pick the one with the smallest third objective value and denote it by $x^{(2)}$. Set \mathcal{R}_2 contains all solutions with a smaller third objective value (encircled points in Figure 2c). Note that in particular no solution of the gray region is considered anymore. On the other hand, x belongs to \mathcal{R}_2 due to Pareto-optimality.

The set \mathcal{C}_1 contains all solutions from \mathcal{R}_2 that have a smaller first objective value than x (encircled points in the gray area in Figure 2d). Among these solutions $x^{(1)}$ is the one with the smallest second objective value. Set \mathcal{R}_1 contains all solutions from \mathcal{R}_2 with a smaller second objective value (encircled points in Figure 2e). This set still contains x , but no points from the gray region.

In the final iteration $t = 0$ we obtain $\mathcal{C}_0 = \mathcal{R}_1$ since there is no restriction in the construction of \mathcal{C}_0 anymore and $\mathcal{C}_0 \neq \emptyset$ since $x \in \mathcal{R}_1$. Solution $x^{(0)}$ is among the remaining solutions the one with the smallest first objective value (Figure 2f). This solution equals x and is now returned.

Let us now discuss how the draw of V can be divided into two steps such that in the first step enough information is revealed to execute the **Witness** function and such that in the second step there is still enough randomness left to bound the probability that $x^{(0)}$ lies in B . For this let $I \subseteq [n]$ be a set of indices and assume that we know in advance which values the solutions $x^{(0)}, \dots, x^{(d)}$ take at these indices, i.e., assume that we know $a^{(0)} = x^{(0)}|_I, \dots, a^{(d)} = x^{(d)}|_I$ before executing the **Witness** function. Then we can reconstruct $x^{(0)}, \dots, x^{(d)}$ without having to reveal the entire matrix V . This can be done by the following algorithm, which gets as additional parameters the set I and the matrix $A = [a^{(0)}, \dots, a^{(d)}]$.

Witness(V, I, A, B)

- 1: Set $\mathcal{R}_{d+1} = \bigcup_{t'=0}^d \mathcal{S}_I(a^{(t')})$.
- 2: **for** $t = d, d-1, \dots, 0$ **do**
- 3: Set $\mathcal{C}_t = \{z \in \mathcal{R}_{t+1} : V^{1\dots t} z \leq b|_{1\dots t}\} \cap \mathcal{S}_I(a^{(t)})$.
- 4: Set $x^{(t)} = \arg \min \{V^{t+1} z : z \in \mathcal{C}_t\}$.
- 5: Set $\mathcal{R}_t = \{z \in \mathcal{R}_{t+1} : V^{t+1} z < V^{t+1} x^{(t)}\} \cap \bigcup_{t'=0}^{t-1} \mathcal{S}_I(a^{(t')})$.
- 6: **end for**
- 7: **return** $(x^{(0)}, \dots, x^{(d)})$

The additional restriction of the set \mathcal{R}_{d+1} does not change the outcome of the **Witness** function as all solutions $x^{(0)}, \dots, x^{(d)}$ generated by the first **Witness** function are contained in the set \mathcal{R}_{d+1} defined in line 1 of the second **Witness** function. Similarly one can argue that the additional restrictions in lines 3 and 5 do not change the outcome of the algorithm because all solutions $x^{(t)}$ generated by the first **Witness** function satisfy the restrictions that are made in the second **Witness** function. Hence, if $a^{(0)} = x^{(0)}|_I, \dots, a^{(d)} = x^{(d)}|_I$, then both **Witness** functions generate the same $x^{(0)}$.

We will now discuss how much information about V needs to be revealed in order to execute the second **Witness** function, assuming that the additional parameters I and A are given. We assume

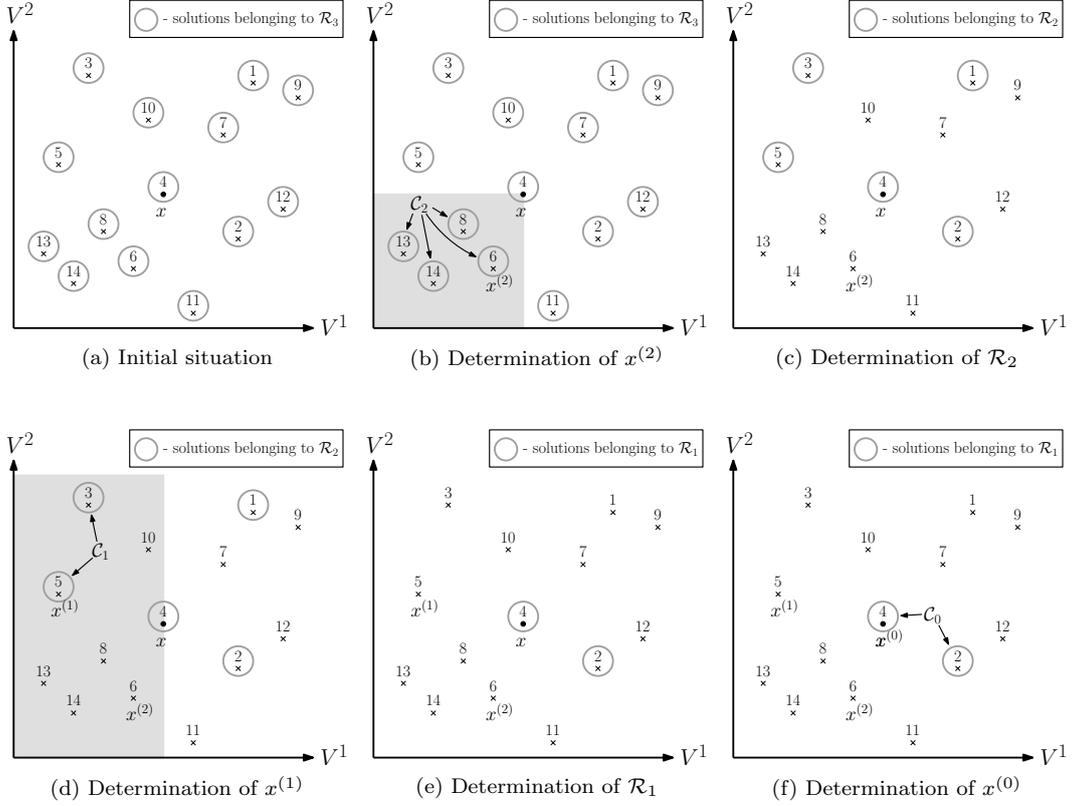


Figure 2: Execution of the **Witness** function for three objectives

that the coefficients V_i^t are revealed for every $t \in [d]$ and $i \notin I$. For the remaining coefficients only certain linear combinations need to be known in order to be able to execute the **Witness** function. By carefully looking at the **Witness** function, one can deduce that for $t \in [d]$ only the following linear combinations need to be known:

$$V^t|_I \cdot x^{(t)}|_I, \dots, V^t|_I \cdot x^{(d)}|_I, \\ V^t|_I \cdot (x^{(t-1)} - x^{(0)})|_I, \dots, V^t|_I \cdot (x^{(t-1)} - x^{(t-2)})|_I.$$

These terms can be viewed as linear combinations of the random variables V_i^t , $t \in [d]$, $i \in I$, with coefficients from $\{-\mathcal{K}, \dots, \mathcal{K}\}$. In addition to the already fixed random variables V_i^t , $t \in [d]$, $i \notin I$, the following d linear combinations determine the position $V^{1 \dots d}x$ of $x = x^{(0)}$:

$$V_I^1 \cdot x^{(0)}|_I, \dots, V_I^d \cdot x^{(0)}|_I.$$

An important observation on which our analysis is based is that if the vectors $x^{(0)}|_I, \dots, x^{(d)}|_I$ are linearly independent, then also all of the above mentioned linear combinations are linearly independent. In particular, the d linear combinations that determine the position of x cannot be expressed by the other linear combinations. Usually, however, it is not possible to find a tuple $I \subseteq [n]$ of indices such that the vectors $x^{(0)}|_I, \dots, x^{(d)}|_I$ are linearly independent. By certain technical modifications of the **Witness** function we will ensure that there always exists such a tuple I with $|I| = d + 1$ and that the last index of I is determined by the other d indices. Since we do not know the tuple I and the matrix A in advance, we apply a union bound over all valid

choices for these parameters, which yields a factor of $(\mathcal{K} + 1)^{(d+1)^2} n^d \leq 2^{(d+1)^2} \mathcal{K}^{(d+1)^2} n^d$ in the bound for the probability that there exists a Pareto-optimal solution in B .

Röglin and Teng [18] observed that the linear independence of the linear combinations implies that even if the linear combinations needed to execute the **Witness** function are revealed in the first step, there is still enough randomness in the second step to prove an upper bound on the probability that $V^{1\dots d}x$ lies in a fixed ε -box B that is proportional to ε^d . The bound proven in [18] is, however, not strong enough to improve Moitra and O'Donnell's result [14] because the dependence on ϕ is in the order of $\Theta(\phi^{d^2})$ which is worse than the dependence of $\Theta(\phi^{d(d+1)/2})$ proven by Moitra and O'Donnell. We show that for quasiconcave density functions the dependence in [18] can be improved significantly to $\Theta(\phi^d)$, which yields the improved bound of $O(n^{2d}\phi^d)$ in Theorem 3 for the binary case.

Higher Moments The analysis of higher moments is based on running the **Witness** function multiple times. Let us fix a c -tuple (B_1, \dots, B_c) of ε -boxes. As described above, we bound the probability that all of them contain a Pareto-optimal solution. For this, we run the **Witness** function c times. In this way, we get for every $j \in [c]$ a sequence $x^{(j,0)}, \dots, x^{(j,d)}$ of solutions such that $x^{(j,0)}$ is the unique candidate for a Pareto-optimal solution in B_j .

As above, we would like to execute the c calls of the **Witness** function without having to reveal the entire matrix V . Again if we know for a subset $I \subseteq [n]$ the values that the solutions $x^{(j,t)}$, $j \in [c]$, $t \in [d]$, take at these positions, then we do not need to reveal the coefficients V_i^t with $i \in I$ to be able to execute the calls of the **Witness** function. As in the case of the first moment, it suffices to reveal some linear combinations of these coefficients.

In order to guarantee that these linear combinations are linearly independent of the linear combinations that determine the positions of the solutions $x^{(j,0)}$, $j \in [c]$, we need to coordinate the calls of the **Witness** function. Otherwise it might happen that, for example, the linear combinations revealed for executing the first call of the witness function determine already the position of $x^{(2,0)}$, the candidate for a Pareto-optimal solution in B_2 . Assume that the first call of the **Witness** function returns a sequence $x^{(1,0)}, \dots, x^{(1,d)}$ of solutions and that $I_1 \subseteq [n]$ is a set of indices that satisfies the desired property that $x^{(1,0)}|_{I_1}, \dots, x^{(1,d)}|_{I_1}$ are linearly independent. In order to achieve that all solutions generated in the following calls of the **Witness** function are linearly independent of these linear combinations, we do not start a second independent call of the **Witness** function, but we restrict the set of feasible solutions first. Instead of choosing $x^{(2,0)}, \dots, x^{(2,d)}$ among all solutions from \mathcal{S} , we restrict the set of feasible solutions for the second call of the **Witness** function to $\mathcal{S}' = S_{I_1}(x^{(2,0)})$. Although we do not know $x^{(2,0)}$ in advance, we can assume to know some of its entries due to a technical trick. Essentially, all solutions generated in call r of the **Witness** function have to coincide with $x^{(r,0)}$ in all positions that have been selected in one of the previous calls.

This and some additional tricks allow us to ensure that in the end there is a set $I \subseteq [n]$ with $|I| \leq (d+1)c$ such that all vectors $x^{(j,t)}|_I$, $j \in [c]$, $t \in [d]$ are linearly independent. Then we can again use the bound proven in [18] to bound the probability that $V^{1\dots d}x^{(j,0)} \in B_j$ simultaneously for every $j \in [c]$ from above by a term proportional to $\varepsilon^{cd}\phi^{cd^2}$. With our improved bound for quasiconcave density functions, we obtain a bound proportional to $\varepsilon^{cd}\phi^{cd}$. Together with a union bound over all valid choices for I and the values $x^{(j,t)}|_I$, $j \in [c]$, $t \in [d]$, we obtain a bound of $k \cdot \mathcal{K}^{c^2(d+1)^2 + cd^2} n^{cd}\phi^{cd}\varepsilon^c d$ on the probability that all candidates $x^{(j,0)}$ lie in their corresponding ε -boxes for the constant $k = 2^{c^2(d+1)^2 + cd^2 + cd} \cdot (cd(d+1))^{cd^2}$ that is hidden in the O -notation. Together with the bound of $O((n\mathcal{K})^{cd}/\varepsilon^{cd})$ for the number of c -tuples (B_1, \dots, B_c) this implies Theorem 4 as the exponent of \mathcal{K} is $c^2(d+1)^2 + cd^2 + cd \leq (c+1)^2(d+1)^2$.

Zero-preserving Perturbations If we use the same **Witness** function as above also for zero-preserving perturbations, then it can happen that there is a Pareto-optimal solution x in the ε -box B that does not coincide with the solution $x^{(0)}$ returned by the **Witness** function. This problem

occurs, for example, if $V^d \cdot x^{(d-1)} = V^d \cdot x^{(0)}$, which we cannot exclude if we allow zero-preserving perturbations. We recommend to visualize this case for $d = 2$. On the other hand if we knew in advance that $V^d \cdot x^{(d-1)} = V^d \cdot x^{(0)}$, then we could bound the probability of $V^d x^{(0)} \in (b_d, b_d + \varepsilon]$ already after the solution $x^{(d-1)}$ has been generated. Hence, if we were only interested in bounding this probability, we could terminate the **Witness** function already after $x^{(d-1)}$ has been generated. Instead of terminating the **Witness** function at this point entirely, we keep in mind that $V^d \cdot x^{(0)}$ has already been determined and we restart the **Witness** function with the remaining objective functions only.

Let us make this a bit more precise. As long as the solutions $x^{(t)}$ generated by the **Witness** function differ in all objective functions from x , we execute the **Witness** function without any modification. Only if a solution $x^{(t)}$ is generated that agrees with x in some objective functions, we deviate from the original **Witness** function. Let $K \subseteq [d]$ denote the objective functions in which $x^{(t)}$ coincides with x . At this point we can bound the probability that $V^t \cdot x \in (b_t, b_t + \varepsilon]$ simultaneously for all $t \in K$. In order to also deal with the other objectives $t \notin K$, we restart the **Witness** function. In this restart, we ignore all objective functions in K and we execute the **Witness** function as if only objectives $t \notin K$ were present. Additionally we restrict in the restart the set of feasible solutions to those that coincide in the objectives $t \in K$ with x , i.e., to $\{y \in \mathcal{S} : V^t \cdot y = V^t \cdot x \text{ for all } t \in K\}$. With similar techniques as in the analysis of higher moments we ensure that different restarts lead to linearly independent linear combinations.

This exploits that every Pareto-optimal solution x is also Pareto-optimal with respect to only the objective functions V^t with $t \notin K$ if the set \mathcal{S} is restricted to solutions that agree with x in all objective functions V^t with $t \in K$. This property guarantees that whenever the **Witness** function is restarted, x is still a Pareto-optimal solution with respect to the restricted solution set and the remaining objective functions.

It can happen that we have to restart the **Witness** function d times before a unique candidate for a Pareto-optimal solution in B is identified. As in each of these restarts at most d solutions are generated, the total number of solutions that is generated can increase from $d + 1$, as in the case of non-zero-preserving perturbations, to roughly d^2 . The set $I \subseteq [n]$ of indices restricted to which these solutions are linearly independent has a cardinality of at most d^3 . The reason for this increase is that we have to choose more indices to obtain linear independence due to the fixed zeros. Taking a union bound over all valid choices of I , of the values that the generated solutions take at these positions, and of the possibilities when and due to which objectives the restarts occur, yields Theorem 1. This theorem relies again on the result about linearly independent linear combinations of independent random variables from [18] and its improved version for quasiconcave densities that we show in this article.

4 Properties of (Weak) Pareto-optimal Solutions

In this section we will identify the main properties of (weakly) Pareto-optimal solutions that lay the foundation for all variants of the **Witness** function. In the model without zero-preserving perturbations we only need properties of Pareto optima. In the model with zero-preserving perturbations, however, much more work has to be done and there we need the notion of weak Pareto optimality.

We start with an observation that is valid for both Pareto-optimal solutions and weak Pareto-optimal solutions.

Proposition 7. *Let $\mathcal{S} \subseteq \mathbb{R}^n$ be a set of solutions, let $f_1, \dots, f_d: \mathcal{S} \rightarrow \mathbb{R}$ be functions, let x^* be a (weak) Pareto optimum with respect to \mathcal{S} and $\{f_1, \dots, f_d\}$, and let $\mathcal{S}' \subseteq \mathcal{S}$ be a subset of solutions that contains x^* . Then x^* is (weakly) Pareto-optimal with respect to \mathcal{S}' and $\{f_1, \dots, f_d\}$.*

The core idea of the **Witness** functions is given by the following lemma and Corollary 9. It implies that if x is Pareto-optimal with respect to \mathcal{R}_{t+1} and $\{V^1, \dots, V^{t+1}\}$, then x is also Pareto-

optimal with respect to \mathcal{R}_t and $\{V^1, \dots, V^t\}$ (cf. function $\text{Witness}(V, B)$ described in Section 3). Given this as the induction step, it yields that x is Pareto-optimal with respect to \mathcal{R}_1 and $\{V^1\}$. This means that in iteration $t = 0$ we obtain $x^{(0)} = \arg \min \{V^1 z : z \in \mathcal{C}_0\} = x$ because $\mathcal{C}_0 = \mathcal{R}_1$.

Lemma 8. *Let $\mathcal{S} \subseteq \mathbb{R}^n$ be a set of solutions, let $f_1, \dots, f_{t+1} : \mathcal{S} \rightarrow \mathbb{R}$, $t \geq 1$, be functions, and let x^* be a weak Pareto optimum with respect to \mathcal{S} and $\{f_1, \dots, f_{t+1}\}$. We consider the set $\mathcal{C} \subseteq \mathcal{S}$ of solutions that dominate x^* strongly with respect to $\{f_1, \dots, f_t\}$.*

(I) *If $\mathcal{C} = \emptyset$, then x^* is weakly Pareto-optimal with respect to \mathcal{S} and $\{f_1, \dots, f_t\}$.*

(II) *If $\mathcal{C} \neq \emptyset$, then let $\hat{f} = \min_{x \in \mathcal{C}} f_{t+1}(x)$. Then $f_{t+1}(x^*) \leq \hat{f}$. Furthermore, if $f_{t+1}(x^*) < \hat{f}$, then x^* is weakly Pareto-optimal with respect to $\mathcal{R} := \{x \in \mathcal{S} : f_{t+1}(x) < \hat{f}\}$ and $\{f_1, \dots, f_t\}$.*

Proof. Claim (I) holds due to the definition of weak Pareto optimality. Let us consider Claim (II). If the inequality $f_{t+1}(x^*) \leq \hat{f}$ does not hold, then $\hat{x} = \arg \min_{x \in \mathcal{C}} f_{t+1}(x)$ dominates x^* strongly with respect to $\{f_1, \dots, f_{t+1}\}$. This is a contradiction since x^* is weakly Pareto-optimal with respect to \mathcal{S} and $\{f_1, \dots, f_{t+1}\}$.

Now let us show that x^* is weakly Pareto-optimal with respect to \mathcal{R} and $\{f_1, \dots, f_t\}$ if $f_{t+1}(x^*) < \hat{f}$. The condition ensures that $x^* \in \mathcal{R}$. Assume to the contrary that there exists a $y \in \mathcal{R}$ that dominates x^* strongly with respect to $\{f_1, \dots, f_t\}$. Since $\mathcal{R} \subseteq \mathcal{S}$, this implies $y \in \mathcal{C}$. Due to $y \in \mathcal{R}$ we obtain the contradiction $f_{t+1}(y) < \hat{f} \leq f_{t+1}(y)$, where the second inequality follows from the definition of \hat{f} and $y \in \mathcal{C}$. \square

If the functions f_1, \dots, f_t in Lemma 8 are injective, we can also obtain a statement about Pareto optima.

Corollary 9. *Let $\mathcal{S} \subseteq \mathbb{R}^n$ be a set of solutions, let $f_1, \dots, f_{t+1} : \mathcal{S} \rightarrow \mathbb{R}$, $t \geq 1$, be functions, where f_1, \dots, f_t are injective, and let x^* be a Pareto optimum with respect to \mathcal{S} and $\{f_1, \dots, f_{t+1}\}$. We consider the set $\mathcal{C} \subseteq \mathcal{S}$ of solutions that dominate x^* strongly with respect to $\{f_1, \dots, f_t\}$.*

(I) *If $\mathcal{C} = \emptyset$, then x^* is Pareto-optimal with respect to \mathcal{S} and $\{f_1, \dots, f_t\}$.*

(II) *If $\mathcal{C} \neq \emptyset$, then let $\hat{f} = \min_{x \in \mathcal{C}} f_{t+1}(x)$. Then $f_{t+1}(x^*) < \hat{f}$. Furthermore, x^* is Pareto-optimal with respect to $\mathcal{R} := \{x \in \mathcal{S} : f_{t+1}(x) < \hat{f}\}$ and $\{f_1, \dots, f_t\}$.*

Proof. First of all we observe that a solution $y \in \mathcal{S}$ dominates x^* with respect to $\{f_1, \dots, f_t\}$ if and only if y dominates x^* strongly with respect to $\{f_1, \dots, f_t\}$. This is due to the injectivity of the functions f_1, \dots, f_t . Consequently, Claim (I) follows from the definition of Pareto optimality. Let us consider Claim (II). Assume to the contrary that $\hat{f} \leq f_{t+1}(x^*)$. In this case, the solution $\hat{x} = \arg \min_{x \in \mathcal{C}} f_{t+1}(x)$ would dominate x^* with respect to $\{f_1, \dots, f_{t+1}\}$ contradicting the assumption that x^* is Pareto-optimal. Hence, $f_{t+1}(x^*) < \hat{f}$.

Due to Lemma 8, x^* is weakly Pareto-optimal with respect to \mathcal{R} and $\{f_1, \dots, f_t\}$ because every Pareto optimum is also a weak Pareto optimum. As these functions are injective, x^* is even Pareto-optimal with respect to \mathcal{R} and $\{f_1, \dots, f_t\}$. \square

For the model with zero-preserving perturbations we need one more lemma that allows us to handle non-injectivity appropriately.

Lemma 10. *Let $\mathcal{S} \subseteq \mathbb{R}^n$ be a set of solutions, let $f_1, \dots, f_{t+1} : \mathcal{S} \rightarrow \mathbb{R}$, $t \geq 1$, be functions, and let x^* be a Pareto optimum with respect to \mathcal{S} and $\{f_1, \dots, f_{t+1}\}$. Furthermore, let $K \subseteq [t + 1]$ be a tuple of indices and let \mathcal{S}' be a subset of $\{x \in \mathcal{S} : f_k(x) = f_k(x^*) \text{ for all } k \in K\}$. Then x^* is Pareto-optimal with respect to \mathcal{S}' and $\{f_k : k \in [t + 1] \setminus K\}$.*

Proof. Assume to the contrary that x^* is not Pareto-optimal. Then there exists a solution $y \in \mathcal{S}'$ such that y dominates x^* with respect to $\{f_k : k \in [t + 1] \setminus K\}$. Since $f_k(y) = f_k(x^*)$ for all $k \in K$, solution y also dominates x^* with respect to $\{f_1, \dots, f_{t+1}\}$. This contradicts the assumption that x^* is Pareto-optimal. \square

5 Non-zero-preserving Perturbations

5.1 Smoothed Number of Pareto-optimal Solutions

To prove Theorem 3 we assume without loss of generality that $n \geq d+1$ and consider the function given as Algorithm 1 which we call the **Witness** function. It is very similar to the one suggested by Moitra and O'Donnell, but with an additional parameter I . This parameter is a tuple of forbidden indices: it restricts the set of indices we are allowed to choose from. For the analysis of the smoothed number of Pareto-optimal solutions we will set $I = ()$. The parameter becomes important in the next section when we analyze higher moments.

Algorithm 1: Witness(V, x, I)

```

1 set  $I_{d+1} = I$ ;
2 set  $\mathcal{R}_{d+1} = \mathcal{S}_{I_{d+1}}(x)$ ;
3 for  $t = d, d-1, \dots, 0$  do
4   set  $\mathcal{C}_t = \{z \in \mathcal{R}_{t+1} : V^{1\dots t}z < V^{1\dots t}x\}$ ;
5   if  $\mathcal{C}_t \neq \emptyset$  then
6     set  $x^{(t)} = \arg \min \{V^{t+1}z : z \in \mathcal{C}_t\}$ ;
7     if  $t = 0$  then return  $x^{(t)}$ ;
8     set  $i_t = \min \{i \in [n] : x_{i_t}^{(t)} \neq x_{i_t}\}$ ;
9     set  $I_t = I_{t+1} \cup (i_t)$ ;
10    set  $\mathcal{R}_t = \{z \in \mathcal{R}_{t+1} : V^{t+1}z < V^{t+1}x^{(t)}\} \cap \mathcal{S}_{I_t}(x)$ ;
11  else
12    set  $i_t = \min([n] \setminus I_{t+1})$ ;
13    set  $I_t = I_{t+1} \cup (i_t)$ ;
14    set  $x_i^{(t)} = \begin{cases} \min(\{0, \dots, \mathcal{K}\} \setminus \{x_i\}) & \text{if } i = i_t \\ x_i & \text{otherwise} \end{cases}$ ;
15    set  $\mathcal{R}_t = \mathcal{R}_{t+1} \cap \mathcal{S}_{I_t}(x)$ ;
16  end
17 end
18 return  $(\perp, \dots, \perp)$ ;

```

Let us give some remarks about the **Witness** function. Note that $\mathcal{C}_0 = \mathcal{R}_1$ since $V^{1\dots t}z < V^{1\dots t}x$ is no restriction if $t = 0$. In Line 6 ties are broken by taking the lexicographically first solution $x^{(t)}$. For $t \geq 1$ the index i_t in Line 8 exists because $V^1x^{(t)} < V^1x$ which implies $x^{(t)} \neq x$.

Unless stated otherwise, we assume that the following *OK-event* $\text{OK}(V)$ occurs. This event occurs if $|V^k \cdot (y - z)| \geq \varepsilon$ for every $k \in [d]$ and for arbitrary two distinct solutions $y \neq z \in \mathcal{S}$ and if for all $k \in [d]$ there is an index $i \in [n]$ for which $|V_i^k| < 1$. Amongst others, the first property ensures that there is a unique arg min in Line 6 and that the functions V^1, \dots, V^d are injective. The latter property, which holds with probability 1, ensures that $B_V(x) \in \mathbb{B}_\varepsilon$ for all vectors $x \in \{-\mathcal{K}, \dots, \mathcal{K}\}^n$. Later we will see that the OK-event occurs with sufficiently high probability.

Before we start to analyze the **Witness** function, let us discuss the differences between the function **Witness**(V, B) described in Section 3 and the function **Witness**(V, x, I) given as Algorithm 1. As described in Section 3 for the illustrative case $d = 2$, the parameters B and x play exactly the same role if $B = B_V(x)$ assuming that the OK-event holds. As stated earlier, the additional parameter I in the function **Witness**(V, x, I) has no meaning for the analysis of the first moment. To prove Theorem 3, we simply set it to the empty tuple. The case $\mathcal{C}_t \neq \emptyset$ (Line 5) is the interesting case, which is also captured by the function **Witness**(V, B). The case $\mathcal{C}_t = \emptyset$ (Line 11)

is the technical case. Here it is only important that we choose an index i_t that is not an element of I_{t+1} and that the vector $x^{(t)}$ is defined such that $x^{(t)}$ coincides with x in all components $i \in I_{t+1}$ and that it does not coincide with x in component i_t . Note that the vector $x^{(t)}$ as we define it in Line 14 is not necessarily a solution from \mathcal{S} .

In the remainder of this section we only consider the case that x is Pareto-optimal, that I is an arbitrary index tuple with pairwise distinct indices, and that the number $|I|$ of indices contained in I is at most $n - (d + 1)$. This ensures that the indices i_0, \dots, i_d exist.

Lemma 11. *The call $\text{Witness}(V, x, I)$ returns the vector $x^{(0)} = x$.*

Proof. We show the following claim by induction on t .

Claim 1. *For all $t \in [d + 1]$, solution x is Pareto-optimal with respect to \mathcal{R}_t and $\{V^1, \dots, V^t\}$.*

Proof of Claim 1. Note that the functions V^1, \dots, V^d are injective due to the assumption that the OK-event occurs. This allows us to apply Corollary 9. Recalling that $x \in \mathcal{S}_{I'}(x)$ for every index tuple I' , Claim 1 is true for $t = d + 1$ by assumption and due to Proposition 7.

Now let us assume that the claim holds for some value $t + 1$ and consider set \mathcal{C}_t . We distinguish between two cases. If $\mathcal{C}_t = \emptyset$, then $\mathcal{R}_t = \mathcal{R}_{t+1} \cap \mathcal{S}_{I_t}(x)$ and the claim follows from the induction hypothesis, from Corollary 9 (I), and from Proposition 7. If $\mathcal{C}_t \neq \emptyset$, then $\mathcal{R}_t = \mathcal{R}'_t \cap \mathcal{S}_{I_t}(x)$ for $\mathcal{R}'_t = \{z \in \mathcal{R}_{t+1} : V^{t+1}z < V^{t+1}x^{(t)}\}$. Hence, the claim follows from the induction hypothesis, from Corollary 9 (II), and from Proposition 7. \square

In accordance with Claim 1, we obtain for $t = 1$ that x is Pareto-optimal with respect to \mathcal{R}_1 and $\{V^1\}$. In particular, $x \in \mathcal{R}_1 = \mathcal{C}_0 \neq \emptyset$, i.e., $x^{(0)} = x$. This solution will be returned in iteration $t = 0$. \square

At a first glance it seems odd to compute a solution x by calling a function with x as parameter. However, we will see that not all information about x is required to execute the call $\text{Witness}(V, x, I)$. To be a bit more precise, the indices i_1, \dots, i_d and the entries at the positions $i \in I \cup (i_1, \dots, i_d)$ of the vectors $x^{(t)}$ constructed during the execution of the Witness function suffice to simulate the execution of Witness without knowing x completely (see Lemma 14). We will call these information a *certificate* (see Definition 12). For technical reasons we will assume that we also know the entries of the vectors $x^{(t)}$ at position $i_0 = \min([n] \setminus (I \cup (i_1, \dots, i_d)))$ and, for the analysis of higher moments, also at further positions.

For our purpose it is not necessary to know how to obtain the required information about x to reconstruct it. It suffices to know that the set of possible certificates is sufficiently small (see Lemma 17) and that for at least one of them the simulation of the execution of Witness returns x (see Lemma 14). This is one crucial property which will help us to bound the expected number of Pareto-optimal solutions.

Definition 12. Let $x^{(0)}, \dots, x^{(d)}$ be the vectors and I_1 be the index tuple constructed during the call $\text{Witness}(V, x, I)$ and set $i_0 = \min([n] \setminus I_1)$ and $I_0 = I_1 \cup (i_0)$. We call the pair (I_0, A_0) for $A_0 = [x^{(d)}, \dots, x^{(0)}]$ the (V, I) -*certificate* of x . The pair (I_0, A) for $A = A_0|_{I_0}$ is called the *restricted* (V, I) -*certificate* of x . We call a pair (I', A') a *(restricted) I-certificate*, if there exist a realization V such that the OK-event occurs and a Pareto-optimal solution $x \in \mathcal{S}$ such that (I', A') is the (restricted) (V, I) -certificate of x . By $\mathcal{C}(I)$ we denote the set of all restricted I -certificates.

The notation used in this section is summarized in Table 1.

For the analysis of the first moment we only need restricted I -certificates. Our analysis of higher moments requires more knowledge about the vectors $x^{(t)}$ than just the values x_i for $i \in I_0$. The additional indices are, however, depending on further calls of the Witness function which we do not know a priori. This is why we have to define two types of certificates. For the sake of reusability we formulate some statements more general than necessary for this section.

$\mathcal{S} \subseteq \{0, \dots, \mathcal{K}\}^n$	set of feasible solutions
V^1, \dots, V^d	linear objective functions
$V^{d+1}: \mathcal{S} \rightarrow \mathbb{R}$	adversarial objective function
V_1^t, \dots, V_n^t	coefficients of V^t for $t \in [d]$
$f_i^t: [-1, 1] \rightarrow [0, \phi]$	probability density of V_i^t for $t \in [d]$ and $i \in [n]$
$V \in \mathbb{R}^{d \times n}$	matrix of coefficients of V^1, \dots, V^d
$\text{PO}(V)$	number of Pareto-optimal solutions for V
$\text{OK}(V)$	event that $ V^k \cdot (y - z) \geq \varepsilon$ for every $k \in [d]$ and for arbitrary two distinct solutions $y \neq z \in \mathcal{S}$ and that for all $k \in [d]$ there is an index $i \in [n]$ for which $ V_i^k < 1$
\mathbb{B}_ε	set of all ε -boxes having corners b for which $b \in \{-n\mathcal{K}, -n\mathcal{K} + \varepsilon, \dots, n\mathcal{K} - 2\varepsilon, n\mathcal{K} - \varepsilon\}^d$
$B_V(x)$	ε -box B for which $V^{1 \dots d} x \in B$
$x^{(0)}, \dots, x^{(d)}$	vectors constructed during the call of Algorithm 1
$\mathcal{R}_{d+1}, \dots, \mathcal{R}_0$	sets constructed during the call of Algorithm 1
$\mathcal{C}_d, \dots, \mathcal{C}_0$	sets constructed during the call of Algorithm 1
I_1	index tuple constructed during the call of Algorithm 1
i_0	$\min([n] \setminus I_1)$
I_0	$I_1 \cup (i_0)$
(I_0, A_0)	(V, I) -certificate of x where $A_0 = [x^{(d)}, \dots, x^{(0)}]$
(I_0, A)	restricted (V, I) -certificate of x where $A = A_0 _{I_0}$
(I', A')	(restricted) I -certificate, i.e., there exist a realization V such that the OK -event occurs and a Pareto-optimal solution $x \in \mathcal{S}$ such that (I', A') is the (restricted) (V, I) -certificate of x
$\mathcal{C}(I)$	set of all restricted I -certificates
$u \in \{0, \dots, \mathcal{K}\}^n$	shift vector used in Algorithm 2

Table 1: Notation used in Section 5.1

Lemma 13. *Let V be an arbitrary realization for which the OK -event occurs, let x be a Pareto-optimal solution with respect to \mathcal{S} and V , and let (I_0, A) be the restricted (V, I) -certificate of x . Then $I_0 = (j_1, \dots, j_{|I|+d+1})$ consists of pairwise distinct indices and*

$$A = \begin{bmatrix} x_{j_1} & \dots & x_{j_{|I|}} & \overline{x_{j_{|I|+1}}} & * & \dots & * \\ \vdots & & \vdots & x_{j_{|I|+1}} & \ddots & \ddots & \vdots \\ \vdots & & \vdots & \vdots & \ddots & \overline{x_{j_{|I|+d}}} & * \\ x_{j_1} & \dots & x_{j_{|I|}} & x_{j_{|I|+1}} & \dots & x_{j_{|I|+d}} & x_{j_{|I|+d+1}} \end{bmatrix}^T \in \{0, \dots, \mathcal{K}\}^{(|I|+d+1) \times (d+1)},$$

where each $*$ can be an arbitrary value from $\{0, \dots, \mathcal{K}\}$ (different $*$ -entries can represent different values) and where \overline{z} for a value $z \in \{0, \dots, \mathcal{K}\}$ can be an arbitrary value from $\{0, \dots, \mathcal{K}\} \setminus \{z\}$.

Proof. Lemma 11 implies that the last column $(x_{j_1}, \dots, x_{j_{|I|+d+1}})^T$ of A equals $x_{I_0}^{(0)} = x|_{I_0}$. Hence, we just have to consider the first d columns of A . Note that $I = (j_1, \dots, j_{|I|})$ and $j_{|I|+1}, \dots, j_{|I|+d+1} = i_d, \dots, i_0$. The construction of the sets \mathcal{R}_t yields $\mathcal{R}_t \subseteq \mathcal{S}_{I_t}(x)$ (see Lines 2, 10, and 15). Index i_t is always chosen such that $i_t \notin I_{t+1}$: If it is constructed in Line 8, then $x_{i_t}^{(t)} \neq x_{i_t}$. Since in this case we have

$$x^{(t)} \in \mathcal{C}_t \subseteq \mathcal{R}_{t+1} \subseteq \mathcal{S}_{I_{t+1}}(x),$$

index i_t cannot be an element of I_{t+1} . In Line 12, index i_t is explicitly constructed such that $i_t \notin I_{t+1}$. The same argument holds for index i_0 . Hence, the indices of I_0 are pairwise distinct.

Now, consider the column of A corresponding to vector $x^{(t)}$ for $t \in [d]$. If $\mathcal{C}_t = \emptyset$, then the form of the column follows directly from the construction of $x^{(t)}$ in Line 14 and from the fact that the indices of I_0 are pairwise distinct. If $\mathcal{C}_t \neq \emptyset$, then

$$x^{(t)} \in \mathcal{C}_t \subseteq \mathcal{R}_{t+1} \subseteq \mathcal{S}_{I_{t+1}}(x),$$

i.e., $x^{(t)}$ coincides with x in all indices $i \in I_{t+1}$. By the choice of i_t in Line 8 we get $x_{i_t}^{(t)} \in \{0, \dots, \mathcal{K}\} \setminus \{x_{i_t}\}$. This concludes the proof. \square

Let (I_0, A_0) be the (V, I) -certificate of x and let $J \supseteq I_0$ be a tuple of pairwise distinct indices. As mentioned before, our goal is to execute the **Witness** function without revealing the entire matrix V . For this we consider the following variant of the **Witness** function given as Algorithm 2 that uses information about x given by the index tuple J , the matrix $A = A_0|_J$ with columns $a^{(d)}, \dots, a^{(0)}$, a shift vector u and the ε -box $B = B_V(x - u)$ instead of vector x itself. The meaning of the shift vector will become clear when we analyze the probability of certain events. We will see that not all information about V needs to be revealed to execute the new **Witness** function, i.e., we have some randomness left which we can use later. With the choice of the shift vector we can control which information has to be revealed for executing the **Witness** function.

Algorithm 2: $\text{Witness}(V, J, A, B, u)$

```

1 let  $b$  be the corner of  $B$  ;
2 set  $\mathcal{R}_{d+1} = \bigcup_{s=0}^d \mathcal{S}_J(a^{(s)})$  ;
3 for  $t = d, d-1, \dots, 0$  do
4   set  $\mathcal{C}_t = \{z \in \mathcal{R}_{t+1} : V^{1\dots t} \cdot (z - u) \leq b|_{1\dots t}\} \cap \mathcal{S}_J(a^{(t)})$  ;
5   if  $\mathcal{C}_t \neq \emptyset$  then
6     set  $x^{(t)} = \arg \min \{V^{t+1}z : z \in \mathcal{C}_t\}$  ;
7     if  $t = 0$  then return  $x^{(t)}$  ;
8     set  $\mathcal{R}_t = \{z \in \mathcal{R}_{t+1} : V^{t+1}z < V^{t+1}x^{(t)}\} \cap \bigcup_{s=0}^{t-1} \mathcal{S}_J(a^{(s)})$  ;
9   else
10    set  $x^{(t)} = (\perp, \dots, \perp)$  ;
11    set  $\mathcal{R}_t = \mathcal{R}_{t+1} \cap \bigcup_{s=0}^{t-1} \mathcal{S}_J(a^{(s)})$  ;
12  end
13 end
14 return  $x^{(0)}$  ;
```

Lemma 14. *Let (I_0, A_0) be the (V, I) -certificate of x , let $J \supseteq I_0$ be an arbitrary tuple of pairwise distinct indices, let $A = A_0|_J$, let $u \in \{0, \dots, \mathcal{K}\}^n$ be an arbitrary vector, and let $B = B_V(x - u)$. Then the call $\text{Witness}(V, J, A, B, u)$ returns vector x .*

Before we give a formal proof of Lemma 14 we try to give some intuition for it. Instead of considering the whole set \mathcal{S} of solutions we restrict it to vectors that look like the vectors we want to reconstruct in the next iterations, i.e., we intersect the current set with the set $\bigcup_{s=0}^{t-1} \mathcal{S}_J(a^{(s)})$ in iteration t . In this way we only deal with subsets of the original sets, but we do not lose the vectors we want to reconstruct since $J \supseteq I_0$. This restriction to the essential candidates of solutions allows us to execute this variant of the **Witness** function with only partial information about V .

Proof. Let $\mathcal{R}'_t, \mathcal{C}'_t$, and $x'^{(t)}$ denote the sets and vectors constructed during the execution of the call $\text{Witness}(V, J, A, B, u)$ and let $\mathcal{R}_t, \mathcal{C}_t$, and $x^{(t)}$ denote the sets and vectors constructed during the execution of call $\text{Witness}(V, x, I)$. We prove the following claims simultaneously by induction.

Claim 2. $\mathcal{R}'_t \subseteq \mathcal{R}_t$ for all $t \in [d+1]$.

Claim 3. $x^{(t)} = x^{(t)}$ for all $t \in [d]_0$ for which $\mathcal{C}_t \neq \emptyset$.

Claim 4. $x^{(s)} \in \mathcal{R}'_t$ for all $t \in [d+1]$ and all $s \in [t-1]_0$ for which $\mathcal{C}_s \neq \emptyset$.

Proof of Claim 2, Claim 3, and Claim 4. Let us first focus on the shift vector u and compare Line 4 of the first **Witness** function (Algorithm 1) with Line 4 of the second **Witness** function (Algorithm 2). The main difference is that in the first version we have the restriction $V^{1\dots t}z < V^{1\dots t}x$, whereas in the second version we seek for solutions z such that $V^{1\dots t} \cdot (z - u) \leq b|_{1\dots t}$. As b is the corner of the ε -box $B = B_V(x - u)$, those restrictions are equivalent for solutions $z \in \mathcal{S}$ since

$$V^{1\dots t} \cdot (z - u) \leq b|_{1\dots t} \iff V^{1\dots t} \cdot (z - u) < V^{1\dots t} \cdot (x - u) \iff V^{1\dots t}z < V^{1\dots t}x.$$

The first inequality is due to the occurrence of the OK-event.

Now we prove the statements by downward induction over t . Let $t = d+1$. Lemma 13 yields $a^{(s)}|_I = x|_I$ for all $s \in [d]_0$, i.e., $\bigcup_{s=0}^d \mathcal{S}_J(a^{(s)}) \subseteq \mathcal{S}_I(x)$ because $I \subseteq I_0 \subseteq J$. Consequently, $\mathcal{R}'_{d+1} \subseteq \mathcal{R}_{d+1}$ (Claim 2). Consider an arbitrary index $s \in [(d+1)-1]_0$ for which $\mathcal{C}_s \neq \emptyset$. Then

$$x^{(s)} \in \mathcal{C}_s \subseteq \mathcal{R}_{s+1} \subseteq \mathcal{S}$$

(see Line 6) and, thus, $x^{(s)} \in \mathcal{S}_J(a^{(s)})$. Hence, $x^{(s)} \in \mathcal{R}'_{d+1}$ (Claim 4).

For the induction step let $t \leq d$. By the observation above we have

$$\begin{aligned} \mathcal{C}'_t &= \{z \in \mathcal{R}'_{t+1} : V^{1\dots t}z < V^{1\dots t}x\} \cap \mathcal{S}_J(a^{(t)}) \quad \text{and} \\ \mathcal{C}_t &= \{z \in \mathcal{R}_{t+1} : V^{1\dots t}z < V^{1\dots t}x\}. \end{aligned}$$

Since $\mathcal{R}'_{t+1} \subseteq \mathcal{R}_{t+1}$, we obtain $\mathcal{C}'_t \subseteq \mathcal{C}_t$. We first consider the case $\mathcal{C}_t = \emptyset$ which implies $\mathcal{C}'_t = \emptyset$ and $t \geq 1$ in accordance with Lemma 11 since $\mathcal{C}_0 \neq \emptyset$. Then

$$\begin{aligned} \mathcal{R}'_t &= \mathcal{R}'_{t+1} \cap \bigcup_{s=0}^{t-1} \mathcal{S}_J(a^{(s)}) \quad \text{and} \\ \mathcal{R}_t &= \mathcal{R}_{t+1} \cap \mathcal{S}_{I_t}(x). \end{aligned}$$

According to Lemma 13, all vectors $x^{(0)}, \dots, x^{(t-1)}$ coincide with x on the indices $i \in I_t$ as $I_t \subseteq I_0 \subseteq J$. Thus, $\bigcup_{s=0}^{t-1} \mathcal{S}_J(a^{(s)}) \subseteq \mathcal{S}_{I_t}(x)$. As $\mathcal{R}'_{t+1} \subseteq \mathcal{R}_{t+1}$ due to Claim 2 of the induction hypothesis, we obtain $\mathcal{R}'_t \subseteq \mathcal{R}_t$ (Claim 2). For Claim 3 nothing has to be shown here. Let $s \in [t-1]_0$ be an index for which $\mathcal{C}_s \neq \emptyset$. Then $x^{(s)} \in \mathcal{R}'_{t+1}$ by Claim 4 of the induction hypothesis, $x^{(s)} \in \mathcal{S}_J(a^{(s)})$, and consequently $x^{(s)} \in \mathcal{R}'_t$ (Claim 4).

Finally, let us consider the case $\mathcal{C}_t \neq \emptyset$. Claim 4 of the induction hypothesis yields $x^{(t)} \in \mathcal{R}'_{t+1}$. Since $x^{(t)} \in \mathcal{S}_J(a^{(t)})$ and $V^{1\dots t}x^{(t)} < V^{1\dots t}x$, also $x^{(t)} \in \mathcal{C}'_t$ and, thus, $\mathcal{C}'_t \neq \emptyset$. Hence, $x^{(t)} = x^{(t)}$ as $\mathcal{C}'_t \subseteq \mathcal{C}_t$ (Claim 3). The remaining claims have only to be validated if $t \geq 1$. Then

$$\mathcal{R}'_t = \{z \in \mathcal{R}'_{t+1} : V^{t+1}z < V^{t+1}x^{(t)}\} \cap \bigcup_{s=0}^{t-1} \mathcal{S}_J(a^{(s)})$$

because $x^{(t)} = x^{(t)}$, and

$$\mathcal{R}_t = \{z \in \mathcal{R}_{t+1} : V^{t+1}z < V^{t+1}x^{(t)}\} \cap \mathcal{S}_{I_t}(x).$$

With the same argument used for the case $\mathcal{C}_t = \emptyset$ we obtain $\mathcal{R}'_{t+1} \cap \bigcup_{s=0}^{t-1} \mathcal{S}_J(a^{(s)}) \subseteq \mathcal{R}_{t+1} \cap \mathcal{S}_{I_t}(x)$ and, hence, $\mathcal{R}'_t \subseteq \mathcal{R}_t$ (Claim 2). Consider an arbitrary index $s \in [t-1]_0$ for which $\mathcal{C}_s \neq \emptyset$. Then

$$x^{(s)} \in \mathcal{C}_s \subseteq \mathcal{R}_{s+1} \subseteq \mathcal{R}_t.$$

In particular, $V^{t+1}x^{(s)} < V^{t+1}x^{(t)}$ (see Line 10) and, hence, $V^{t+1}x^{(s)} < V^{t+1}x^{(t)}$ because $x^{(t)} = x^{(t)}$. Furthermore, $x^{(s)} \in \mathcal{R}'_{t+1}$ due to the induction hypothesis, Claim 4, and $x^{(s)} \in \mathcal{S}_J(a^{(s)})$. Consequently, $x^{(s)} \in \mathcal{R}'_t$ (Claim 4). \square

With the claims above Lemma 14 follows immediately: Since $x^{(0)} = x$ and $\mathcal{C}_0 \neq \emptyset$ due to Lemma 11, we obtain $x'^{(0)} = x^{(0)}$ (Claim 3). Hence, the call $\mathbf{Witness}(V, J, A, B, u)$ returns $x'^{(0)} = x$. \square

As mentioned earlier, with the shift vector u we control which information of V has to be revealed to execute the call $\mathbf{Witness}(V, J, A, B, u)$. While Lemma 14 holds for every vector u , we have to choose u carefully for our probabilistic analysis to work. We will see that the choice $u^* = u^*(J, A)$, given by

$$u_i^* = \begin{cases} |x_i - 1| & \text{if } i = i_0, \\ x_i & \text{if } i \in J \setminus (i_0), \\ 0 & \text{otherwise,} \end{cases} \quad (1)$$

is appropriate since $x_i - u_i^* = 0$ for all $i \in J \setminus (i_0)$ and $|x_{i_0} - u_{i_0}^*| = 1$ (cf. Lemma 19). Recall that i_0 is the index that has been added to I_1 in the definition of the (V, I) -certificate to obtain I_0 and note that $u^* \in \{0, \dots, \mathcal{K}\}^n$. Moreover, for every index $i \in J$ the value x_i is given in the last column of A (see Lemma 13). Hence, if (I_0, A_0) is the (V, I) -certificate of x , then vector u^* can be defined when a tuple $J \supseteq I_0$ and the matrix $A = A_0|_J$ are known; we do not have to know the solution x itself.

For bounding the number of Pareto-optimal solutions consider the functions $\chi_{I_0, A, B}(V)$ parameterized by an arbitrary restricted I -certificate (I_0, A) , and an arbitrary ε -box $B \in \mathbb{B}_\varepsilon$, defined as follows: $\chi_{I_0, A, B}(V) = 1$ if the call $\mathbf{Witness}(V, I_0, A, B, u^*(I_0, A))$ returns a solution $x' \in \mathcal{S}$ for which $B_V(x' - u^*(I_0, A)) = B$, and $\chi_{I_0, A, B}(V) = 0$ otherwise.

Corollary 15. *Assume that the OK-event occurs. Then the number $PO(V)$ of Pareto-optimal solutions is at most*

$$\sum_{(I_0, A) \in \mathcal{C}(I)} \sum_{B \in \mathbb{B}_\varepsilon} \chi_{I_0, A, B}(V).$$

Proof. Let x be a Pareto-optimal solution, let (I_0, A) be the restricted (V, I) -certificate of x , and let $B = B_V(x - u^*(I_0, A)) \in \mathbb{B}_\varepsilon$. Due to Lemma 14, $\mathbf{Witness}(V, I_0, A, B, u^*(I_0, A))$ returns vector x . Hence, $\chi_{I_0, A, B}(V) = 1$. It remains to show that the assignment $x \mapsto (I_0, A, B)$ given in the previous lines is injective. Otherwise we would count the occurrence of two distinct Pareto-optimal solutions x_1 and x_2 only once in the sum stated in Corollary 15.

Let x_1 and x_2 be distinct Pareto-optimal solutions and let $(I_0^{(1)}, A_1)$ and $(I_0^{(2)}, A_2)$ be the restricted (V, I) -certificates of x_1 and x_2 , respectively. If $(I_0^{(1)}, A_1) \neq (I_0^{(2)}, A_2)$, then x_1 and x_2 are mapped to distinct triplets. Otherwise, $u^*(I_0^{(1)}, A_1) = u^*(I_0^{(2)}, A_2)$ and, hence, $B_V(x_1 - u^*(I_0^{(1)}, A_1)) \neq B_V(x_2 - u^*(I_0^{(2)}, A_2))$ because of the OK-event and $x_1 \neq x_2$. Consequently, also in this case x_1 and x_2 are mapped to distinct triplets. \square

Corollary 15 immediately implies a bound on the expected number of Pareto-optimal solutions.

Corollary 16. *The expected number of Pareto-optimal solutions is bounded by*

$$\mathbf{E}_V[PO(V)] \leq \sum_{(I_0, A) \in \mathcal{C}(I)} \sum_{B \in \mathbb{B}_\varepsilon} \mathbf{Pr}_V[E_{I_0, A, B}] + (\mathcal{K} + 1)^n \cdot \mathbf{Pr}_V[\overline{OK(V)}],$$

where $E_{I_0, A, B}$ denotes the event that the call $\mathbf{Witness}(V, I_0, A, B, u^*(I_0, A))$ returns a vector x' such that $B_V(x' - u^*(I_0, A)) = B$.

Proof. By applying Corollary 15, we obtain

$$\begin{aligned}
& \mathbf{E}_V[\text{PO}(V)] \\
&= \mathbf{E}_V[\text{PO}(V) \mid \text{OK}(V)] \cdot \mathbf{Pr}_V[\text{OK}(V)] + \mathbf{E}_V[\text{PO}(V) \mid \overline{\text{OK}}(V)] \cdot \mathbf{Pr}_V[\overline{\text{OK}}(V)] \\
&\leq \mathbf{E}_V \left[\sum_{(I_0, A) \in \mathcal{C}(I)} \sum_{B \in \mathbb{B}_\varepsilon} \chi_{I_0, A, B}(V) \mid \text{OK}(V) \right] \cdot \mathbf{Pr}_V[\text{OK}(V)] + |\mathcal{S}| \cdot \mathbf{Pr}_V[\overline{\text{OK}}(V)] \\
&\leq \mathbf{E}_V \left[\sum_{(I_0, A) \in \mathcal{C}(I)} \sum_{B \in \mathbb{B}_\varepsilon} \chi_{I_0, A, B}(V) \right] + (\mathcal{K} + 1)^n \cdot \mathbf{Pr}_V[\overline{\text{OK}}(V)] \\
&= \sum_{(I_0, A) \in \mathcal{C}(I)} \sum_{B \in \mathbb{B}_\varepsilon} \mathbf{Pr}_V[E_{I_0, A, B}] + (\mathcal{K} + 1)^n \cdot \mathbf{Pr}_V[\overline{\text{OK}}(V)]. \quad \square
\end{aligned}$$

We will see that the first term of the sum in Corollary 16 can be bounded independently of ε and that the limit of the second term tends to 0 for $\varepsilon \rightarrow 0$. First of all, we analyze the size of the restricted certificate space.

Lemma 17. *The size of the restricted certificate space $\mathcal{C}(I)$ for $I = ()$ is bounded by*

$$|\mathcal{C}(I)| \leq (\mathcal{K} + 1)^{(d+1)^2} n^d.$$

Proof. Exactly d indices i_1, \dots, i_d are created during the execution of the call $\text{Witness}(V, x, I)$ if the OK-event occurs and if x is Pareto-optimal with respect to V . The index i_0 is determined deterministically depending on the indices i_1, \dots, i_d . Matrix A of every restricted I -certificate (I_0, A) is a $(d+1) \times (d+1)$ -matrix with entries from $\{0, \dots, \mathcal{K}\}$. Hence, the number of possible restricted I -certificates is bounded by $(\mathcal{K} + 1)^{(d+1)^2} n^d$. \square

Let us now fix an arbitrary I -certificate (I_0, A_0) , a tuple $J \supseteq I_0$, and an ε -box $B \in \mathbb{B}_\varepsilon$. We want to analyze the probability $\mathbf{Pr}_V[E_{J, A, B}]$ where $A = A_0|_J$. By V_J and $V_{\overline{J}}$ we denote the part of the matrix $V^{1 \dots d}$ that belongs to the indices $i \in J$ and to the indices $i \notin J$, respectively. We apply the principle of deferred decisions and assume that $V_{\overline{J}}$ is fixed as well, i.e., we will only exploit the randomness of V_J .

As motivated above, the call $\text{Witness}(V, J, A, B, u)$ can be executed without the full knowledge of V_J . To formalize this, we introduce matrices Q_k that describe the linear combinations of V_J^k that suffice to be known:

$$Q_k = [p^{(d)}, \dots, p^{(k)}, p^{(k-2)} - p^{(k-1)}, \dots, p^{(0)} - p^{(k-1)}] \in \{-\mathcal{K}, \dots, \mathcal{K}\}^{|J| \times d} \quad (2)$$

for $p^{(t)} = p^{(t)}(J, A, u) = a^{(t)} - u|_J$ where $a^{(t)}$ are the columns of matrix $A = [a^{(d)}, \dots, a^{(0)}]$ and $t \in [d]_0$. Note that the matrices $Q_k = Q_k(J, A, u)$ depend on the pair (J, A) and on the vector u .

Lemma 18. *Let $u \in \{0, \dots, \mathcal{K}\}^n$ be an arbitrary shift vector and let U and W be two realizations of V such that $U_{\overline{J}} = W_{\overline{J}}$ and $U_J^k \cdot q = W_J^k \cdot q$ for all indices $k \in [d]$ and all columns q of the matrix $Q_k(J, A, u)$. Then the calls $\text{Witness}(U, J, A, B, u)$ and $\text{Witness}(W, J, A, B, u)$ return the same result.*

Lemma 18 states that for different realizations U_J and W_J of V_J the modified Witness function outputs the same result. Actually, in the proof we will even see that the complete execution of both calls is identical. This means that solution x is already determined if these realizations are known. However, there is still randomness left in the objective values $V^1 x, \dots, V^d x$ which allows us to bound the probability that x falls into box B (see Corollary 21).

Proof. We fix an index $k \in [d]$ and analyze which information of V_J^k is required for the execution of the call $\text{Witness}(V, J, A, B, u)$. For the execution of Line 4 we need to know $V^k \cdot (z - u)$ for solutions $z \in \mathcal{S}_J(a^{(t)})$ in all iterations $t \geq k$. Since we assume V_J^k to be known, this means that

$$V_J^k \cdot (z|_J - u|_J) = V_J^k \cdot (a^{(t)} - u|_J) = V_J^k \cdot p^{(t)}$$

must be revealed. For $t \geq k$ vector $p^{(t)}$ is a column of Q_k . The execution of Line 6 does not require further information about V_J^k : The only iteration where we might need information about V_J^k is iteration $t = k - 1$. However, as $\mathcal{C}_t \subseteq \mathcal{S}_J(a^{(t)})$, we obtain

$$x^{(t)} = \arg \min \{V^{t+1}z : z \in \mathcal{C}_t\} = \arg \min \{V_J^{t+1}z|_J : z \in \mathcal{C}_t\}$$

because all solutions $z \in \mathcal{C}_t$ agree on the entries with indices $i \in J$. Since $V_J^{t+1} = V_J^k$ is known, $x^{(t)}$ can be determined without any further information. Note that this does not imply that $V^{t+1}x^{(t)}$ is already specified.

It remains to consider Line 8. Only in iteration $t = k - 1$ we need information about V^k . In that iteration it suffices to know $V_J^k \cdot (z|_J - x^{(t)}|_J)$ for every solution $z \in \bigcup_{s=0}^{t-1} \mathcal{S}_J(a^{(s)})$. Hence, for $z \in \mathcal{S}_J(a^{(s)})$, $s \in [t-1]_0 = [k-2]_0$, the linear combinations

$$V_J^k \cdot (z|_J - x^{(t)}|_J) = V_J^k \cdot ((a^{(s)} - u|_J) - (a^{(k-1)} - u|_J)) = V_J^k \cdot (p^{(s)} - p^{(k-1)})$$

must be revealed. For $s \in [k-2]_0$, vector $p^{(s)} - p^{(k-1)}$ is a column of Q_k .

As U and W agree on all necessary information, both calls return the same result. \square

We will now see why $u^* = u^*(J, A)$ defined in Equation (1) is a good shift vector.

Lemma 19. *Let $Q = [\hat{p}^{(d)}, \dots, \hat{p}^{(0)}]$ where $\hat{p}^{(t)} = p^{(t)}(J, A, u^*(J, A))|_{I_0}$. Then*

$$|Q| = \begin{bmatrix} 0 & \dots & 0 & + & * & \dots & * \\ \vdots & & \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & & \vdots & \vdots & \ddots & + & * \\ 0 & \dots & 0 & 0 & \dots & 0 & 1 \end{bmatrix}^T \in \{0, \dots, \mathcal{K}\}^{(|I|+d+1) \times (d+1)},$$

where $|Q|$ denotes the matrix Q' for which $q'_{ij} = |q_{ij}|$. Each $'*'$ -entry can be an arbitrary value from $\{0, \dots, \mathcal{K}\}$ and each $'+'$ -entry can be an arbitrary value from $\{1, \dots, \mathcal{K}\}$. Different $'*'$ -entries as well as different $'+'$ -entries can represent different values.

Proof. Let $I_0 = (j_1, \dots, j_{|I|+d+1})$, i.e., $i_0 = j_{|I|+d+1}$. According to Lemma 13 and the construction of vector u^* in Equation (1) we obtain

$$Q = \begin{bmatrix} x_{j_1} & \dots & \dots & x_{j_1} \\ \vdots & & & \vdots \\ x_{j_{|I|}} & \dots & \dots & x_{j_{|I|}} \\ \overline{x_{j_{|I|+1}}} & x_{j_{|I|+1}} & \dots & x_{j_{|I|+1}} \\ * & \ddots & \ddots & \vdots \\ \vdots & \ddots & \overline{x_{j_{|I|+d}}} & x_{j_{|I|+d}} \\ * & \dots & * & x_{j_{|I|+d+1}} \end{bmatrix} - \begin{bmatrix} x_{j_1} & \dots & x_{j_1} \\ \vdots & & \vdots \\ x_{j_{|I|}} & \dots & x_{j_{|I|}} \\ x_{j_{|I|+1}} & \dots & x_{j_{|I|+1}} \\ \vdots & & \vdots \\ x_{j_{|I|+d}} & \dots & x_{j_{|I|+d}} \\ |x_{j_{|I|+d+1}} - 1| & \dots & |x_{j_{|I|+d+1}} - 1| \end{bmatrix}.$$

The claim follows since $|a - b| \leq \mathcal{K}$, $\bar{a} - a \neq 0$, and $|a - |a - 1|| = 1$ for all $a, b \in \{0, \dots, \mathcal{K}\}$. \square

Lemma 20. For all $k \in [d]$ the columns of the matrix $Q_k(J, A, u^*(J, A))$ and the vector $p^{(0)}$ are linearly independent.

Proof. Let $\hat{p}^{(t)} = p^{(t)}|_{I_0}$ for all $t \in [d]_0$. It suffices to show that the columns of the submatrix $\hat{Q}_k = Q_k|_{I_0}$ and the vector $\hat{p}^{(0)}$ are linearly independent. Consider the matrix $Q = [\hat{p}^{(d)}, \dots, \hat{p}^{(0)}]$. Due to Lemma 19 the last $d + 1$ rows of Q form a lower triangular matrix and the entries on the principal diagonal are from the set $\{-\mathcal{K}, \dots, \mathcal{K}\} \setminus \{0\}$. Consequently, the vectors $\hat{p}^{(t)}$ are linearly independent. As these vectors are the same as the columns of matrix \hat{Q}_1 plus vector $\hat{p}^{(0)}$ (see Equation 2), the claim holds for $k = 1$. Now let $k \geq 2$. We consider an arbitrary linear combination of the columns of \hat{Q}_k and the vector $\hat{p}^{(0)}$ and show that it is 0 if and only if all coefficients are 0.

$$\begin{aligned} 0 &= \sum_{t=k}^d \lambda_t \cdot \hat{p}^{(t)} + \sum_{t=0}^{k-2} \lambda_t \cdot (\hat{p}^{(t)} - \hat{p}^{(k-1)}) + \mu \cdot \hat{p}^{(0)} \\ &= \sum_{t=k}^d \lambda_t \cdot \hat{p}^{(t)} + \sum_{t=1}^{k-2} \lambda_t \cdot \hat{p}^{(t)} - \left(\sum_{t=0}^{k-2} \lambda_t \right) \cdot \hat{p}^{(k-1)} + (\lambda_0 + \mu) \cdot \hat{p}^{(0)}. \end{aligned}$$

As the vectors $\hat{p}^{(t)}$ are linearly independent, we first get $\lambda_t = 0$ for $t \in [d] \setminus \{k-1\}$, which yields $\lambda_0 = 0$ due to $\sum_{t=0}^{k-2} \lambda_t = 0$ and, finally, $\mu = 0$ because of $\lambda_0 + \mu = 0$. This concludes the proof. \square

Corollary 21. Let $\gamma = d(d+1)$. For an arbitrary restricted I -certificate (I_0, A) the probability of the event $E_{I_0, A, B}$ is bounded by

$$\Pr_V [E_{I_0, A, B}] \leq (2\gamma\mathcal{K})^{\gamma-d} \phi^\gamma \varepsilon^d$$

and by

$$\Pr_V [E_{I_0, A, B}] \leq 2^d (\gamma\mathcal{K})^{\gamma-d} \phi^d \varepsilon^d$$

if all densities are quasiconcave.

Proof. Event $E_{I_0, A, B}$ occurs if the output of the call $\text{Witness}(V, I_0, A, B, u^*(I_0, A))$ is a vector x' for which $B_V(x' - u^*(I_0, A)) = B$. We apply the principle of deferred decisions and assume that $V|_{\overline{I_0}}$ is fixed arbitrarily. Now let us further assume that the linear combinations of $V_{I_0}^k$ given by the columns of matrix $Q_k = Q_k(I_0, A, u^*(I_0, A))$ are known for all $k \in [d]$. This means that for some fixed values we consider all realizations of V for which the linear combinations of $V_{I_0}^k$ given by the columns of Q_k equal these values. In accordance with Lemma 18, vector x' is therefore already determined, i.e., it is the same for all realizations of V that are still under consideration.

The equality $B_V(x' - u^*(I_0, A)) = B$ holds if and only if

$$V^k \cdot (x' - u^*(I_0, A)) = V_{I_0}^k \cdot (x' - u^*(I_0, A))|_{\overline{I_0}} + V_{I_0}^k \cdot (x' - u^*(I_0, A))|_{I_0} \in (b_k, b_k + \varepsilon]$$

holds for all $k \in [d]$, where $b = (b_1, \dots, b_d)$ is the corner of B . Since

$$(x' - u^*(I_0, A))|_{I_0} = a^{(0)} - u^*(I_0, A)|_{I_0} = p^{(0)}$$

for the vector $p^{(0)} = p^{(0)}(I_0, A, u^*(I_0, A))$, this is equivalent to the event that

$$V_{I_0}^k \cdot p^{(0)} \in (b_k, b_k + \varepsilon] - V_{I_0}^k \cdot (x' - u^*(I_0, A))|_{\overline{I_0}} =: C_k,$$

where C_k is an interval of length ε depending on x' and hence on the linear combinations of V_{I_0} given by the matrices Q_k . By C we denote the d -dimensional hypercube $C = \prod_{k=1}^d C_k$ with side length ε defined by the intervals C_k .

For all $k \in [d]$ let $Q'_k \in \{-\mathcal{K}, \dots, \mathcal{K}\}^{|I_0| \times (d+1)}$ be the matrix consisting of the columns of Q_k and the vector $p^{(0)}$. These matrices form the diagonal blocks of the matrix

$$Q' = \begin{bmatrix} Q'_1 & \mathbb{O} & \dots & \mathbb{O} \\ \mathbb{O} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbb{O} \\ \mathbb{O} & \dots & \mathbb{O} & Q'_d \end{bmatrix} \in \{-\mathcal{K}, \dots, \mathcal{K}\}^{d \cdot (|I|+d+1) \times d \cdot (d+1)}.$$

Lemma 20, applied for $J = I_0$, implies that matrix Q' has full rank. We permute the columns of Q' to obtain a matrix Q whose last d columns belong to the last column of one of the matrices Q'_k . This means that the last d columns of Q' are $(p^{(0)}, 0^{|I_0|}, \dots, 0^{|I_0|}), \dots, (0^{|I_0|}, \dots, 0^{|I_0|}, p^{(0)})$. Let the rows of Q be labeled by $Q_{j_1,1}, \dots, Q_{j_m,1}, \dots, Q_{j_1,d}, \dots, Q_{j_m,d}$ assuming that $I_0 = (j_1, \dots, j_m)$. We introduce random variables $X_{j,k} = V_j^k$, $j \in I_0$, $k \in [d]$, labeled in the same fashion as the rows of Q . Event $E_{I_0,A,B}$ holds if and only if the d linear combinations of the variables $X_{j,k}$ given by the last d columns of Q fall into the d -dimensional hypercube C depending on the linear combinations of the variables $X_{j,k}$ given by the remaining columns of Q . The claim follows by applying Theorem 40 for the matrix Q^T and $k = d$ and due to the fact that the number of columns of Q is $\gamma = d \cdot (d+1)$. Hence,

$$\Pr_V [E_{I_0,A,B}] \leq (2\gamma\mathcal{K})^{\gamma-d} \phi^\gamma \varepsilon^d$$

in general and

$$\Pr_V [E_{I_0,A,B}] \leq 2^d (\gamma\mathcal{K})^{\gamma-d} \phi^d \varepsilon^d$$

if all densities are quasiconcave. The different bounds for general densities and quasiconcave densities come solely from Theorem 40. \square

Proof of Theorem 3. We begin the proof by showing that the OK-event is likely to happen. For all indices $t \in [d]$ and all solutions $x \neq y \in \mathcal{S}$ the probability that $|V^t x - V^t y| \leq \varepsilon$ is bounded by $2\phi\varepsilon$. To see this, choose one index $i \in [n]$ such that $x_i \neq y_i$ and apply the principle of deferred decisions by fixing all coefficients V_j^t for $j \neq i$. Then the value V_i^t must fall into an interval of length $2\varepsilon/|x_i - y_i| \leq 2\varepsilon$. The probability for this is bounded from above by $2\varepsilon\phi$. A union bound over all indices $t \in [d]$ and over all pairs $(x, y) \in \mathcal{S} \times \mathcal{S}$ for which $x \neq y$ yields $\Pr_V [\overline{\text{OK}}(V)] \leq 2(\mathcal{K}+1)^{2n} d\phi\varepsilon$.

Let $\gamma = d \cdot (d+1)$. We set

$$s = \begin{cases} (2\gamma\mathcal{K})^{\gamma-d} \phi^\gamma & \text{for general density functions,} \\ 2^d (\gamma\mathcal{K})^{\gamma-d} \phi^d & \text{for quasiconcave density functions.} \end{cases}$$

With $I = ()$ we obtain

$$\begin{aligned} \mathbf{E}_V [\text{PO}(V)] &\leq \sum_{(I_0,A) \in \mathcal{C}(I)} \sum_{B \in \mathbb{B}_\varepsilon} \Pr_V [E_{I_0,A,B}] + (\mathcal{K}+1)^n \cdot \Pr_V [\overline{\text{OK}}(V)] \\ &\leq \sum_{(I_0,A) \in \mathcal{C}(I)} \sum_{B \in \mathbb{B}_\varepsilon} s \cdot \varepsilon^d + (\mathcal{K}+1)^n \cdot 2(\mathcal{K}+1)^{2n} d\phi\varepsilon \\ &= |\mathcal{C}(I)| \cdot |\mathbb{B}_\varepsilon| \cdot s \cdot \varepsilon^d + 2(\mathcal{K}+1)^{3n} d\phi\varepsilon \\ &\leq (\mathcal{K}+1)^{(d+1)^2} n^d \cdot \left(\frac{2n\mathcal{K}}{\varepsilon}\right)^d \cdot s \cdot \varepsilon^d + 2(\mathcal{K}+1)^{3n} d\phi\varepsilon \\ &= 2^d (\mathcal{K}+1)^{(d+1)^2} \mathcal{K}^d n^{2d} \cdot s + 2(\mathcal{K}+1)^{3n} d\phi\varepsilon. \end{aligned}$$

The first inequality is due to Corollary 16. The second inequality is due to Corollary 21. The third inequality stems from Lemma 17. Since this bound is true for every $\varepsilon > 0$ for which $1/\varepsilon$ is integral,

it also holds for the limit $\varepsilon \rightarrow 0$. Hence, we obtain

$$\mathbf{E}_V[\text{PO}(V)] \leq 2^d(\mathcal{K} + 1)^{(d+1)^2} \mathcal{K}^d n^{2d} \cdot s.$$

Substituting s and γ by their definitions yields

$$\begin{aligned} \mathbf{E}_V[\text{PO}(V)] &\leq 2^d(\mathcal{K} + 1)^{(d+1)^2} \mathcal{K}^d n^{2d} \cdot (2d(d+1)\mathcal{K})^{d(d+1)-d} \phi^{d(d+1)} \\ &= \mathcal{K}^{2(d+1)^2} \cdot O(n^{2d} \phi^{d(d+1)}) \end{aligned}$$

for general densities and

$$\begin{aligned} \mathbf{E}_V[\text{PO}(V)] &\leq 2^d(\mathcal{K} + 1)^{(d+1)^2} \mathcal{K}^d n^{2d} 2^d (d(d+1)\mathcal{K})^{d(d+1)-d} \phi^d \\ &= \mathcal{K}^{2(d+1)^2} \cdot O(n^{2d} \phi^d) \end{aligned}$$

for quasiconcave densities. □

5.2 Higher Moments

The basic idea behind our analysis of higher moments is the following: If the OK-event occurs, then we can count the c^{th} power of the number $\text{PO}(V)$ of Pareto-optimal solutions by counting all c -tuples (B_1, \dots, B_c) of ε -boxes where each ε -box B_i contains a Pareto-optimal solution x_i . We can bound this value as follows: First, call $\text{Witness}(V, x_1, ())$ to obtain a vector x'_1 and consider the index tuple $I_0^{(1)}$ that contains all indices created in this call and one additional index. In the second step, call $\text{Witness}(V, x_2, I_0^{(1)})$ to obtain a vector x'_2 and consider the tuple $I_0^{(2)}$ consisting of the indices of $I_0^{(1)}$, the indices created in this call, and one additional index. Now, call $\text{Witness}(V, x_3, I_0^{(2)})$ and so on. For the call $\text{Witness}(V, x, I)$ to be well-defined, in Section 5.1 we assumed $|I| \leq n - (d+1)$. Consequently, here we have to ensure that $|I_0^{(c-1)}| \leq n - (d+1)$, i.e., $n \geq c \cdot (d+1)$. We can assume this for fixed integers c and d because all of our results are presented in O -notation.

If (x_1, \dots, x_c) is a tuple of Pareto-optimal solutions with $V^{1 \dots d} x_i \in B_i$ for $i \in [c]$, then $(x'_1, \dots, x'_c) = (x_1, \dots, x_c)$ due to Lemma 11. As in the analysis of the first moment, we use the variant of the **Witness** function that uses certificates of the vectors x_ℓ instead of the vectors itself to simulate the calls. Hence we can reuse several statements of Section 5.1.

Let us remark that for bounding the c^{th} moment of the smoothed number of Pareto-optimal solutions we also have to consider c -tuples (B_1, \dots, B_c) of ε -boxes for which $B_k = B_\ell$ for some indices $k < \ell$. This might seem critical as both boxes contain the same Pareto-optimal solution $x_k = x_\ell$ (if such a solution exists) which could cause problems due to dependencies. We resolve this problem by using different shift vectors u_k and u_ℓ for reconstructing the vectors x_k and x_ℓ with the **Witness** function.

Unless stated otherwise, let V be a realization such that the OK-event $\text{OK}(V)$ occurs and fix arbitrary solutions $x_1, \dots, x_c \in \mathcal{S}$ with $V^{1 \dots d} x_i \in B_i$ for $i \in [c]$ that are Pareto-optimal with respect to V .

Definition 22. Let $I_0^{(0)} = ()$ and let $(I_0^{(\ell)}, A_0^{(\ell)})$ be the $(V, I_0^{(\ell-1)})$ -certificate of x_ℓ defined in Definition 12, $\ell = 1, \dots, c$. We call the pair (I, A) for $I = I_0^{(c)}$, $A = (A^{(1)}, \dots, A^{(c)})$, and $A^{(\ell)} = A_0^{(\ell)}|_I$, the *(restricted) V -certificate* of (x_1, \dots, x_c) . We call a pair (I_0', A') a *c -certificate*, if there is a realization V such that the OK-event occurs and if there are Pareto-optimal solutions $x_1, \dots, x_c \in \mathcal{S}$ such that (I_0', A') is the V -certificate of (x_1, \dots, x_c) . By \mathcal{C}_c we denote the set of all c -certificates.

Note that $I_0^{(0)} \subseteq \dots \subseteq I_0^{(c)}$ and $|I_0^{(\ell)}| = |I_0^{(\ell-1)}| + d + 1$ for $\ell \in [c]$.

We now consider the functions $\chi_{I,A,\vec{B}}(V)$, parameterized by an arbitrary c -certificate $(I, A) \in \mathcal{C}_c$ and a vector $\vec{B} = (B_1, \dots, B_c) \in \mathbb{B}_\varepsilon^c$ of ε -boxes, which is defined as follows: $\chi_{I,A,\vec{B}}(V) = 1$ if for all $\ell \in [c]$ the call $\text{Witness}(V, I, A^{(\ell)}, B_\ell, u^*(I, A^{(\ell)}))$ returns a solutions x'_ℓ such that $B_V(x'_\ell - u^*(I, A^{(\ell)})) = B_\ell$, and $\chi_{I,A,\vec{B}}(V) = 0$ otherwise. Recall that the vector $u^* = u^*(I, A^{(\ell)})$ is defined in Equation 1.

Corollary 23. *Assume that the OK-event occurs. Then the c^{th} power of the number $PO(V)$ of Pareto-optimal solutions is at most*

$$\sum_{(I,A) \in \mathcal{C}_c} \sum_{\vec{B} \in \mathbb{B}_\varepsilon^c} \chi_{I,A,\vec{B}}(V).$$

Proof. The c^{th} power of the number $PO(V)$ of Pareto-optimal solutions equals the number of c -tuples (x_1, \dots, x_c) of Pareto-optimal solutions. Let (x_1, \dots, x_c) be such a c -tuple, let (I, A) be the V -certificate of (x_1, \dots, x_c) , and let $B_\ell = B_V(x_\ell - u^*(I, A^{(\ell)})) \in \mathbb{B}_\varepsilon$. Due to Lemma 14, $\text{Witness}(V, I, A^{(\ell)}, B_\ell, u^*(I, A^{(\ell)}))$ returns vector x_ℓ for all $\ell \in [c]$. Hence, $\chi_{I,A,\vec{B}}(V) = 1$ for $\vec{B} = (B_1, \dots, B_c)$. As in the proof of Corollary 15 we have to show that this assignment $(x_1, \dots, x_c) \mapsto (I, A, \vec{B})$ is injective.

Let (x_1, \dots, x_c) and (y_1, \dots, y_c) be distinct c -tuples of Pareto-optimal solutions, i.e., there is an index $\ell \in [c]$ such that $x_\ell \neq y_\ell$, and let (I_1, A_1) and (I_2, A_2) be their V -certificates. If $(I_1, A_1) \neq (I_2, A_2)$, then both tuples are mapped to distinct triplets. Otherwise, $u^*(I_1, A_1^{(\ell)}) = u^*(I_2, A_2^{(\ell)})$ and, thus, $B_V(x_\ell - u^*(I_1, A_1^{(\ell)})) \neq B_V(y_\ell - u^*(I_2, A_2^{(\ell)}))$ because of the OK-event and $x_\ell \neq y_\ell$. Consequently, also in this case (x_1, \dots, x_c) and (y_1, \dots, y_c) are mapped to distinct triplets. \square

Corollary 23 immediately implies a bound on the c^{th} moment of the number of Pareto-optimal solutions.

Corollary 24. *The c^{th} moment of the number of Pareto-optimal solutions is bounded by*

$$\mathbf{E}_V[PO^c(V)] \leq \sum_{(I,A) \in \mathcal{C}_c} \sum_{\vec{B} \in \mathbb{B}_\varepsilon^c} \Pr_V \left[E_{I,A,\vec{B}} \right] + (\mathcal{K} + 1)^{cn} \cdot \Pr_V \left[\overline{OK(V)} \right],$$

where $E_{I,A,\vec{B}}$ denotes the event that $\chi_{I,A,\vec{B}}(V) = 1$.

We omit the proof since it is exactly the same as the one of Corollary 16.

Lemma 25. *The size of the certificate space is bounded by*

$$|\mathcal{C}_c| \leq (\mathcal{K} + 1)^{c^2(d+1)^2} n^{cd}.$$

Proof. Let (I, A) be an arbitrary c -certificate. Each matrix $A^{(\ell)}$ is a $|I| \times (d+1)$ -matrix with entries from $\{0, \dots, \mathcal{K}\}$. The tuple I can be written as

$$I = (i_d^{(1)}, \dots, i_0^{(1)}, \dots, i_d^{(c)}, \dots, i_0^{(c)}),$$

created by c successive calls of the Witness function, where the indices $i_0^{(\ell)}$ are chosen deterministically in Definition 12. Since $|I| = c \cdot (d+1)$ the claim follows. \square

Corollary 26. *Let $\gamma = cd(d+1)$. For an arbitrary c -certificate (I, A) and an arbitrary vector $\vec{B} \in \mathbb{B}_\varepsilon^c$ of ε -boxes the probability of the event $E_{I,A,\vec{B}}$ is bounded by*

$$\Pr_V \left[E_{I,A,\vec{B}} \right] \leq (2\gamma\mathcal{K})^{\gamma-cd} \phi^{\gamma} \varepsilon^{cd}$$

and by

$$\Pr_V \left[E_{I,A,\vec{B}} \right] \leq 2^{cd} (\gamma\mathcal{K})^{\gamma-cd} \phi^{cd} \varepsilon^{cd}$$

if all densities are quasiconcave.

Proof. For $k \in [d]$ and $\ell \in [c]$ consider the matrices $Q_k(I, A^{(\ell)}, u_\ell^*)$ for $u_\ell^* = u^*(I, A^{(\ell)})$ defined in Equation (2). Due to Lemma 18 the output of the call $\mathbf{Witness}(V, I, A^{(\ell)}, B_\ell, u_\ell^*)$ is determined if V_I and the linear combinations $V_I^k \cdot q$ for all indices $k \in [d]$ and all columns q of the matrix $Q_k^{(\ell)} = Q_k(I, A^{(\ell)}, u_\ell^*)$ are given. With the same argument as in the proof of Corollary 26 event $E_{I, A, \bar{B}}$ occurs if and only if $V_I \cdot [p^{(\ell, 1)}, \dots, p^{(\ell, d)}]$ falls into some d -dimensional hypercube C_ℓ with side length ε depending on the linear combinations $V_I \cdot Q_k^{(\ell)}$. In this notation, $p^{(\ell, t)}$ is short for $p^{(t)}(I, A^{(\ell)}, u_\ell^*)$.

Now, consider the matrix

$$Q'_k = [Q_k^{(1)}, p^{(1, k)}, \dots, Q_k^{(c)}, p^{(c, k)}] \in \{-\mathcal{K}, \dots, \mathcal{K}\}^{|I| \times c \cdot (d+1)}.$$

Note that $|I| = c \cdot (d+1) = \gamma/d$. Due to Lemma 19, Q'_k is a lower block triangular matrix, due to Lemma 20 the columns of $[Q_k^{(\ell)}, p^{(\ell, k)}]$ are linearly independent. Hence, matrix Q'_k is an invertible matrix and the same holds for the block diagonal matrix

$$Q' = \begin{bmatrix} Q'_1 & \mathbb{O} & \dots & \mathbb{O} \\ \mathbb{O} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbb{O} \\ \mathbb{O} & \dots & \mathbb{O} & Q'_d \end{bmatrix} \in \{-\mathcal{K}, \dots, \mathcal{K}\}^{\gamma \times \gamma}.$$

We permute the columns of Q' to obtain a matrix Q where the last cd columns belong to the columns $p^{(1, 1)}, \dots, p^{(1, d)}, \dots, p^{(c, 1)}, \dots, p^{(c, d)}$. We assume the rows of Q to be labeled by $Q_{j_1, 1}, \dots, Q_{j_m, 1}, \dots, Q_{j_1, d}, \dots, Q_{j_m, d}$ where $I = (j_1, \dots, j_m)$ and introduce random variables $X_{j, k} = V_j^k$, $j \in I$, $k \in [d]$, indexed the same way as the rows of Q . Event $E_{I, A, \bar{B}}$ holds if and only if the cd linear combinations of the variables $X_{j, k}$ given by the last cd columns of Q fall into the cd -dimensional hypercube $C = \prod_{\ell=1}^c C_\ell$ with side length ε depending on the linear combinations of the variables $X_{j, k}$ given by the remaining columns of Q . The claim follows by applying Theorem 40 for the matrix Q^T and $k = cd$ and due to the fact that the number of columns of Q is γ . \square

Proof of Theorem 4. In the proof of Theorem 3 we showed that the probability that the OK-event does not hold is bounded by $2(\mathcal{K} + 1)^{2n} d\phi\varepsilon$. Let $\gamma = cd(d+1)$. We set

$$s = \begin{cases} (2\gamma\mathcal{K})^{\gamma - cd} \phi^\gamma & \text{for general density functions,} \\ 2^{cd} (\gamma\mathcal{K})^{\gamma - cd} \phi^{cd} & \text{for quasiconcave density functions.} \end{cases}$$

Then we obtain

$$\begin{aligned} \mathbf{E}_V[\text{PO}^c(V)] &\leq \sum_{(I, A) \in \mathcal{C}_c} \sum_{\bar{B} \in \mathbb{B}_\varepsilon^c} \mathbf{Pr}_V[E_{I, A, \bar{B}}] + (\mathcal{K} + 1)^{cn} \cdot \mathbf{Pr}_V[\overline{\text{OK}(V)}] \\ &\leq \sum_{(I, A) \in \mathcal{C}_c} \sum_{\bar{B} \in \mathbb{B}_\varepsilon^c} s \cdot \varepsilon^{cd} + (\mathcal{K} + 1)^{cn} \cdot 2(\mathcal{K} + 1)^{2n} d\phi\varepsilon \\ &= |\mathcal{C}_c| \cdot |\mathbb{B}_\varepsilon^c| \cdot s \cdot \varepsilon^{cd} + 2(\mathcal{K} + 1)^{(c+2)n} d\phi\varepsilon \\ &\leq (\mathcal{K} + 1)^{c^2(d+1)^2} n^{cd} \cdot \left(\frac{2n\mathcal{K}}{\varepsilon}\right)^{cd} \cdot s \cdot \varepsilon^{cd} + 2(\mathcal{K} + 1)^{(c+2)n} d\phi\varepsilon \\ &= 2^{cd} (\mathcal{K} + 1)^{c^2(d+1)^2} \mathcal{K}^{cd} n^{2cd} \cdot s + 2(\mathcal{K} + 1)^{(c+2)n} d\phi\varepsilon. \end{aligned}$$

The first inequality is due to Corollary 24. The second inequality is due to Corollary 26. The third inequality stems from Lemma 25. Since this bound is true for every $\varepsilon > 0$ for which $1/\varepsilon$ is integral, it also holds for the limit $\varepsilon \rightarrow 0$. Hence, we obtain

$$\mathbf{E}_V[\text{PO}^c(V)] \leq 2^{cd} (\mathcal{K} + 1)^{c^2(d+1)^2} \mathcal{K}^{cd} n^{2cd} \cdot s.$$

Substituting s and γ by their definitions yields

$$\begin{aligned}\mathbf{E}_V[\text{PO}^c(V)] &\leq 2^{cd}(\mathcal{K}+1)^{c^2(d+1)^2} \mathcal{K}^{cd} n^{2cd} \cdot (2cd(d+1)\mathcal{K})^{cd(d+1)-cd} \phi^{cd(d+1)} \\ &= \mathcal{K}^{(c+1)^2(d+1)^2} \cdot O((n^{2d} \phi^{d(d+1)})^c)\end{aligned}$$

for general densities and

$$\begin{aligned}\mathbf{E}_V[\text{PO}^c(V)] &\leq 2^{cd}(\mathcal{K}+1)^{c^2(d+1)^2} \mathcal{K}^{cd} n^{2cd} \cdot 2^{cd} (cd(d+1)\mathcal{K})^{cd(d+1)-cd} \phi^{cd} \\ &= \mathcal{K}^{(c+1)^2(d+1)^2} \cdot O((n^{2d} \phi^d)^c)\end{aligned}$$

for quasiconcave densities. □

The proof of Theorem 4 yields that $\mathbf{E}_V[\text{PO}^c(V)] \leq s_c$ for

$$s_c := 2^{c(d+1)^2} (cd(d+1))^{cd^2} (\mathcal{K}+1)^{(c+1)^2(d+1)^2} n^{2cd} \phi^{c\beta},$$

where

$$\beta = \begin{cases} d(d+1) & \text{for general density functions,} \\ d & \text{for quasiconcave density functions.} \end{cases}$$

With the following Corollary we bound the probability that $\text{PO}(V)$ exceeds a certain multiple of s_1 . We obtain a significantly better concentration bound than the one we would obtain by applying Markov's inequality for the first moment.

Corollary 27. *The probability that the number of Pareto-optimal solutions is at least $\lambda \cdot s_1$ for some $\lambda \geq 1$ is bounded by*

$$\Pr_V[\text{PO}(V) \geq \lambda \cdot s_1] \leq \left(\frac{1}{\lambda}\right)^{\frac{1}{2} \cdot \left\lfloor \frac{\log_{\mathcal{K}+1} \lambda}{4(d+1)^2} \right\rfloor}.$$

Proof. Let c^* be the real for which $(\mathcal{K}+1)^{2c^*(d+1)^2} = \lambda^{1/2}$, i.e.,

$$c^* = \frac{\log_{\mathcal{K}+1} \lambda}{4(d+1)^2}.$$

Observing that $c \leq 2^c \leq (\mathcal{K}+1)^c$ for all $c \in \mathbb{R}$ and setting $c = \lfloor c^* \rfloor$ yields

$$\begin{aligned}\Pr_V[\text{PO}(V) \geq \lambda \cdot s_1] &= \Pr_V[\text{PO}^c(V) \geq \lambda^c \cdot s_1^c] = \Pr_V\left[\text{PO}^c(V) \geq \frac{\lambda^c \cdot s_1^c}{\mathbf{E}_V[\text{PO}^c(V)]} \cdot \mathbf{E}_V[\text{PO}^c(V)]\right] \\ &\leq \frac{\mathbf{E}_V[\text{PO}^c(V)]}{\lambda^c \cdot s_1^c} \leq \frac{s_c}{\lambda^c \cdot s_1^c} = \frac{2^{c(d+1)^2} (cd(d+1))^{cd^2} (\mathcal{K}+1)^{(c+1)^2(d+1)^2} n^{2cd} \phi^{c\beta}}{\lambda^c \cdot 2^{c(d+1)^2} (d(d+1))^{cd^2} (\mathcal{K}+1)^{4c(d+1)^2} n^{2cd} \phi^{c\beta}} \\ &= \frac{c^{cd^2} (\mathcal{K}+1)^{(c-1)^2(d+1)^2}}{\lambda^c} \leq \left(\frac{c^{(d+1)^2} (\mathcal{K}+1)^{c(d+1)^2}}{\lambda}\right)^c \\ &\leq \left(\frac{(\mathcal{K}+1)^{c(d+1)^2} (\mathcal{K}+1)^{c(d+1)^2}}{\lambda}\right)^c \leq \left(\frac{(\mathcal{K}+1)^{2c^*(d+1)^2}}{\lambda}\right)^c \\ &= \left(\frac{1}{\lambda}\right)^{\frac{c}{2}} = \left(\frac{1}{\lambda}\right)^{\frac{1}{2} \cdot \left\lfloor \frac{\log_{\mathcal{K}+1} \lambda}{4(d+1)^2} \right\rfloor}.\end{aligned}$$

The first inequality is Markov's inequality. The second inequality only holds if $c \geq 1$. However, for $c = 0$ the inequality $\Pr_V[\text{PO}(V) \geq \lambda \cdot s_1] \leq \lambda^{-c/2}$ is trivially true. □

6 Zero-preserving Perturbations

Our analysis of Theorem 1 holds for instances with the following property: There exists a partition (I_1, \dots, I_d) of $[n]$ such that, for all $t \in [d]$ and for all $i \in [n]$, the coefficient V_i^t is not deterministically set to 0 if and only if $i \in I_t$. This means that exactly n of the $d \cdot n$ coefficients are perturbed and that the value $V^t x$ only depends on the entries x_i of x for which $i \in I_t$. All other objective functions do not depend on these entries. Furthermore, we require $|I_t| \geq (d+1)^3$ for all $t \in [d]$.

With the next two lemmas we show that if Theorem 1 holds for instances that have the form described above, then it also holds for all other instances with a slightly larger constant that is hidden in the O -notation.

Lemma 28. *Without loss of generality in each objective function except for the adversarial one there are more than $(d+1)^3$ perturbed coefficients, i.e., coefficients that are not deterministically set to 0.*

Proof. For an index $k \in [d]$ let P_k be the tuple of indices i for which V_i^k is a perturbed coefficient, i.e., a coefficient which is not set to 0 deterministically. Let K be the tuple of indices k for which $|P_k| \leq (d+1)^3$, let $P = \bigcup_{k \in K} P_k$, and consider the decomposition of \mathcal{S} into subsets of solutions $\mathcal{S}_v = \{x \in \mathcal{S} : x|_P = v\}$, $v \in \{0, \dots, \mathcal{K}\}^{|P|}$. Let $x \in \mathcal{S}_v$ be an arbitrary solution. If x is Pareto-optimal with respect to \mathcal{S} and $\{V^1, \dots, V^{d+1}\}$, then x is also Pareto-optimal with respect to \mathcal{S}_v and $\{V^k : k \in [d+1] \setminus K\}$ due to Lemma 10. As all remaining objective functions V^k , $k \in [d+1] \setminus K$, have more than $(d+1)^3$ perturbed coefficients, the instance with these objective functions and \mathcal{S}_v as set of feasible solutions has the desired form and we can apply Theorem 1 for each of these instances. Since we now have $(\mathcal{K}+1)^{|P|} \leq (2\mathcal{K})^{|K| \cdot (d+1)^3}$ instances, each having $d - |K|$ linear and one adversarial objective, we can bound the number of Pareto-optimal solutions by

$$(2\mathcal{K})^{|K| \cdot (d+1)^3} \cdot \mathcal{K}^{(d-|K|+1)^5} \cdot O\left(n^{\alpha(d-|K|)} \cdot \phi^{\beta(d-|K|)}\right),$$

where α and β denote the exponents of n and ϕ in the bound stated in Theorem 1. These exponents depend on the number d of non-adversarial objectives and whether the densities are quasiconcave or not. Since they are monotonically increasing, which particularly implies $\alpha(d-|K|) \leq \alpha(d)$ and $\beta(d-|K|) \leq \beta(d)$, we can bound the number of Pareto-optima simply by

$$(2\mathcal{K})^{|K| \cdot (d+1)^3} \cdot \mathcal{K}^{(d-|K|+1)^5} \cdot O\left(n^{\alpha(d)} \cdot \phi^{\beta(d)}\right).$$

Hence, it suffices to show that

$$\mathcal{K}^{|K| \cdot (d+1)^3 + (d-|K|+1)^5} \leq \mathcal{K}^{(d+1)^5},$$

as the additional factor $2^{|K| \cdot (d+1)^3} \leq 2^{(d+1)^3 d}$ can be hidden in the O -notation. This inequality is equivalent to showing that $b \cdot a^3 + (a-b)^5 \leq a^5$ for $b = |K|$ and $a = d+1$. Note that $0 \leq b = |K| \leq d = a-1$. By a chain of equivalences we obtain

$$\begin{aligned} b \cdot a^3 + (a-b)^5 \leq a^5 &\iff a^3 \leq \frac{1}{b} \cdot (a^5 - (a^5 - 5a^4b + 10a^3b^2 - 10a^2b^3 + 5ab^4 - b^5)) \\ &\iff a^3 \leq 5a^4 - 10a^3b + 10a^2b^2 - 5ab^3 + b^4 \\ &= 5a \cdot (a^3 - 2a^2b + 2ab^2 - b^3) + b^4 \\ &= 5a(a-b) \cdot (a^2 - ab + b^2) + b^4 =: f(a, b). \end{aligned}$$

Applying the inequality $ab \leq \frac{(a+b)^2}{4}$ yields

$$f(a, b) \geq 5a(a-b) \cdot \left(a^2 - \frac{(a+b)^2}{4} + b^2\right) \geq 5a \cdot \left(\frac{a^2}{2} + \frac{(a-b)^2}{4} + \frac{b^2}{2}\right) \geq \frac{5}{2}a^3 \geq a^3.$$

This concludes the proof. \square

Lemma 29. *Without loss of generality for every $i \in [n]$ exactly one of the coefficients V_i^1, \dots, V_i^d is perturbed, whereas the others are deterministically set to 0.*

Let us remark that we can transform every instance that has not the form stated in Lemma 29 into an instance with this form. We will show that this transformation does not increase the size of the Pareto-set for any realization of the coefficients. Hence, the bound from Theorem 1 that applies for the modified instance also applies for the original instance. However, our transformation increases the dimension of the set \mathcal{S} from n to $d \cdot n$. Hence, we lose a factor of $d^{d^3+d^2+d}$ in the bound which we can hide in the O -notation since we have to apply this transformation only once.

Proof. We first show how to decrease the number of indices i for which V_i^1, \dots, V_i^d is perturbed to at most one. For this, let

$$\mathcal{S}' = \{(x, x, \dots, x) : x \in \mathcal{S}\} \subseteq \{0, \dots, \mathcal{K}\}^{dn}$$

be the set of feasible solutions that contains for every $x \in \mathcal{S}$ the solution $x^d \in \{0, \dots, \mathcal{K}\}^{dn}$ that consists of d copies of x . For $k \in [d]$ we define a linear objective function $W^k: \mathcal{S}' \rightarrow \mathbb{R}$ in which all coefficients W_i^k with $i \notin I_k := \{(k-1)n+1, \dots, kn\}$ are deterministically set to 0. The coefficients $W_{(k-1)n+1}^k, \dots, W_{kn}^k$ are chosen as the coefficients V_1^k, \dots, V_n^k , i.e., either randomly according to a density f_i^k or 0 deterministically. The objective function W^{d+1} maps every solution $x^d \in \mathcal{S}'$ to $V^{d+1}(x)$. The instance consisting of \mathcal{S}' and the objective functions W^1, \dots, W^{d+1} has the desired property that every variable appears in at most one of the objective functions W^1, \dots, W^d and it has the same smoothed number of Pareto-optimal solutions as the instance consisting of \mathcal{S} and the objective functions V^1, \dots, V^{d+1} . For every $i \in [dn]$ for which none of the coefficients W_i^1, \dots, W_i^d is perturbed we can eliminate the corresponding variable from \mathcal{S}' .

This shows that every ϕ -smooth instance with $\mathcal{S} \subseteq \{0, \dots, \mathcal{K}\}^n$ can be transformed into another ϕ -smooth instance with $\mathcal{S} \subseteq \{0, \dots, \mathcal{K}\}^\ell$ with $\ell \leq dn$ in which every variable appears in exactly one objective function and that has the same smoothed number of Pareto-optimal solutions. As the bound proven in Theorem 1 depends polynomially on the number of variables, we lose only a constant factor (with respect to n , ϕ , and \mathcal{K}) by going from $\mathcal{S} \subseteq \{0, \dots, \mathcal{K}\}^n$ to $\mathcal{S}' \subseteq \{0, \dots, \mathcal{K}\}^{dn}$. This constant is hidden in the O -notation. \square

In the remainder of this chapter we focus on instances having the structure described in Lemma 28 and Lemma 29. Then (P_1, \dots, P_d) is a partition of $[n]$, where P_t denotes the tuple of indices i for which V_i^t is perturbed.

We consider the variant of the **Witness** function given as Algorithm 3, referred to as the **Witness₀** function, which gets as parameters besides the usual V , x , and I , a set $K \subseteq [d]$ of indices of objective functions and a call number $r \in \mathbb{N}$. In a call of the **Witness₀** function only the adversarial objective function V^{d+1} and the objective functions V^t with $t \in K$ are considered. The set of solutions is restricted to solutions that agree with x in all positions P_k with $k \notin K$. Additionally, as in the **Witness** function, only solutions are considered that agree with x in all positions $i \in I$. By the right choice of I , we can avoid choosing an index multiple times in different calls of the **Witness₀** function. The parameter r simply corresponds to the number of the current call of the **Witness₀** function. The **Witness₀** function always returns some subset of \mathcal{S} .

Let us give some remarks about the **Witness₀** function. As a convention we set $\bigcap_{k \in \emptyset} \mathcal{S}_{P_k}(x) = \mathcal{S}$ (cf. Line 3). This is only important in the case where $K = [d]$, i.e., in the first call.

As in the **Witness** function, if iteration $t = 0$ is reached in a certain call r (this does not have to be the case), then we obtain $\mathcal{C}_0^{(r)} = \mathcal{R}_1^{(r)}$ since $V^{k_1 \dots k_t} z < V^{k_1 \dots k_t} x$ (see Line 6) is no restriction for $t = 0$. For the definition of $x^{(r,t)}$ in Line 8, ties are broken by taking the lexicographically first solution. Though we did the same in the **Witness** function, it is much more important here. In the model without zero-preserving perturbations the functions V^1, \dots, V^d are injective with probability 1. If this is the case, then no ties have to be broken. In the model with zero-preserving

Algorithm 3: $\text{Witness}_0(V, x, K, r, I)$

```
1 let  $d_r$  be the number of components of  $K$  and let  $K$  be of the form  $K = (k_1, \dots, k_{d_r})$  ;
2 set  $k_{d_r+1} = d + 1$  ;
3 set  $\mathcal{R}_{d_r+1}^{(r)} = \mathcal{S}_I(x) \cap \bigcap_{k \in [d] \setminus K} \mathcal{S}_{P_k}(x)$  ;
4 if  $d_r = 0$  then return  $\mathcal{R}_{d_r+1}^{(r)}$  ;
5 for  $t = d_r, d_r - 1, \dots, 0$  do
6   set  $\mathcal{C}_t^{(r)} = \{z \in \mathcal{R}_{t+1}^{(r)} : V^{k_1 \dots k_t} z < V^{k_1 \dots k_t} x\}$  ;
7   if  $\mathcal{C}_t^{(r)} \neq \emptyset$  then
8     set  $x^{(r,t)} = \arg \min \{V^{k_{t+1}} z : z \in \mathcal{C}_t^{(r)}\}$  ;
9     let  $K_{\text{eq}} \subseteq K$  be the tuple of indices  $k$  for which  $x^{(r,t)}|_{P_k} = x|_{P_k}$  ;
10    set  $K_{\text{neq}} = K \setminus K_{\text{eq}}$  ;
11    for  $k \in K$  do
12      if  $k \in K_{\text{eq}}$  then
13        set  $r_k = r$  ;
14      else
15        set  $i_k = \min \{i \in P_k : x_i^{(r,t)} \neq x_i\}$  ;
16         $I \leftarrow I \cup (i_k)$  ;
17      end
18    end
19    if  $K_{\text{eq}} = ()$  then
20      set  $\mathcal{R}_t^{(r)} = \{z \in \mathcal{R}_{t+1}^{(r)} : V^{k_{t+1}} z < V^{k_{t+1}} x^{(r,t)}\} \cap \mathcal{S}_I(x)$  ;
21    else
22      set  $t_r = t$  ;
23      return  $\text{Witness}_0(V, x, K_{\text{neq}}, r + 1, I)$  ;
24    end
25  else
26    for  $k \in K$  do
27      set  $i_k = \min(P_k \setminus I)$  ;
28       $I \leftarrow I \cup (i_k)$  ;
29    end
30    set  $x_i^{(r,t)} = \begin{cases} \min(\{0, \dots, \mathcal{K}\} \setminus \{x_i\}) & \text{if } i \in \{i_{k_1}, \dots, i_{k_{d_r}}\} \\ x_i & \text{otherwise} \end{cases}$  ;
31    set  $\mathcal{R}_t^{(r)} = \mathcal{R}_{t+1}^{(r)} \cap \mathcal{S}_I(x)$  ;
32  end
33 end
34 return  $\emptyset$  ;
```

perturbations, the functions V^1, \dots, V^d can be non-injective with probability 1: If there are two distinct solutions $x, y \in \mathcal{S}$ for which $x|_{P_k} = y|_{P_k}$, then $V^k x = V^k y$.

The index r_k defined in Line 13 is the number of the last call in which the objective function V^k has been considered. The index t_r defined in Line 22 is the number of the iteration in call number r of Witness_0 in which the next recursive call of Witness_0 was made. We will see that, if the last call of the Witness_0 function is the call with number $r^* + 1$, then $r_k \in [r^*]$ for each $k \in [d]$ and there is at least one index $k \in [d]$ for which $r_k = r^*$, i.e., the objective function V^k has been considered until the end. Furthermore, the indices t_r are defined for $r = 1, \dots, r^*$. For the simulation of the Witness function information about the solutions $x^{(t)}$ and the indices i_t are required. For the

simulation of the Witness_0 function we additionally need the values r_k and t_r to know when to make a new recursive call (in iteration $t = t_r$ in the call with number r) and which objectives to consider (objective V^k will be considered in the call with number r if and only if $r \leq r_k$).

In Line 15 it is always possible to find an index $i \in P_k$ on which the current vector $x^{(r,t)}$ and x disagree because this line is only reached if $k \in K_{\text{neq}}$, i.e., if $x^{(r,t)}|_{P_k} \neq x|_{P_k}$. In order for Line 27 to be feasible, we have to guarantee that $P_k \setminus I \neq ()$. This follows since we assumed $|P_k| > (d+1)^3 > d(d+1)$ in accordance with Lemma 28 and because there are at most d calls of Witness_0 with non-empty K with at most $d+1$ iterations each, and in each iteration at most one index from P_k is added to I . Note that it would be more precise to introduce the notation $i_k^{(r,t)}$ rather than i_k (cf. Line 15 and Line 27). Furthermore, we could also write $I_k^{(r,t)}$ instead of I . For the sake of readability we decided to drop these additional indices and refer to index i_k and tuple I of iteration t of call r in our proofs.

Before we analyze the Witness_0 function, let us discuss similarities and differences to the Witness function. The initial calls $\text{Witness}_0(V, x, [d], 1, I)$ and $\text{Witness}(V, x, I)$ are very similar. All objectives V^1, \dots, V^{d+1} are considered. Furthermore, $d_1 = d$ and $\mathcal{R}_{d+1}^{(1)} = \mathcal{S}_I(x)$. Line 4 can be ignored in this call since $d_1 = d \geq 1$. Also the loop of the Witness_0 function is very similar to the loop of the Witness function. The sets $\mathcal{C}_t^{(r)}$ and $\mathcal{R}_t^{(r)}$ and the solution $x^{(r,t)}$ are defined the same way as the sets \mathcal{C}_t and \mathcal{R}_t and the solution $x^{(t)}$ in the Witness function.

However, there are two main differences to the Witness function in the body of the loop that are due to the two additional issues we have to deal with when considering zero-preserving perturbations. First it is possible that $V^k x^{(r,t)} = V^k x$ for some of the indices k . This happens if $x^{(r,t)}|_{P_k} = x|_{P_k}$ (otherwise, it happens with probability 0) and is a fundamental issue. If we would proceed running the loop as we do it in the Witness function, then we might lose the crucial property that the function returns $\{x\}$ if x is Pareto-optimal. The tuple K_{eq} contains the problematic indices k for which $x^{(r,t)}|_{P_k} = x|_{P_k}$. If $K_{\text{eq}} = ()$, then we proceed more or less as we did in the Witness function (see Line 15 and Line 20). As discussed above, the case $K_{\text{eq}} \neq ()$ has to be treated differently. In this case we make use of Lemma 10 which implies that, if x is Pareto-optimal, then it is also Pareto-optimal with respect to $\bigcap_{k \in K_{\text{eq}}} \mathcal{S}_{P_k}(x)$ and $\{V^k : k \in K_{\text{neq}} \cup (d+1)\}$ (cf. Line 23 of the current call and Line 3 of the next call).

The second difference due to another issue with zero-preserving perturbations can be sketched as follows. In each iteration t of the Witness function one index i_t is chosen. Since in the model without zero-preserving perturbations all coefficients are perturbed, we can then exploit the randomness in the coefficients $V_{i_t}^1, \dots, V_{i_t}^d$. In the model with zero-preserving perturbations under the assumption given by Lemma 29, for each index $i \in [n]$ exactly one of the coefficients V_i^1, \dots, V_i^d is perturbed while the others are 0. Hence, we choose one index $i_k \in P_k$ for each objective V^k to ensure that we have one perturbed coefficient $V_{i_k}^k$ per objective. These indices i_k are chosen only for $k \in K_{\text{neq}}$ (see Line 15). This is due to the fact that for our analysis to work we need the additional property that $x_{i_k}^{(r,t)} \neq x_{i_k}$ which is impossible for $k \in K_{\text{eq}}$ by the definition of K_{eq} and the requirement $i_k \in P_k$. However, as from now on we do not consider the objectives V^k for $k \in K_{\text{eq}}$ anymore, we do not have to choose indices i_k for $k \in K_{\text{eq}}$.

In the remainder of this section we only consider the case that x is Pareto-optimal. Unless stated otherwise, we assume that the following OK_0 -event $OK_0(V)$ occurs. This event occurs if $|V^k \cdot (y - z)| \geq \varepsilon$ for every $k \in [d]$ and for every two solutions $y, z \in \mathcal{S}$ for which $y|_{P_k} \neq z|_{P_k}$. We will later see that the OK_0 -event occurs with sufficiently high probability.

Lemma 30. *The call $\text{Witness}_0(V, x, [d], 1, ())$ returns the set $\{x^{(r^*, t_{r^*})}\} = \{x\}$, where $r^* = \max\{r_1, \dots, r_d\}$.*

Lemma 30 was also stated in the conference version ([7], Lemma 25) but Claim 1 of the proof was not correct. Here, we rely on the concept of weak Pareto-optimality (see Definition 6) and its properties (Lemma 8) since we cannot guarantee x to be Pareto-optimal in every iteration.

However, the Pareto-optimality holds at the beginning of every call to the Witness_0 function.

Proof. Let us consider an arbitrary call $\text{Witness}_0(V, x, K, r, I)$ for $K \neq ()$. First we show the following claim by induction on t .

Claim 5. *In every iteration t that is reached, x is weakly Pareto-optimal with respect to $\mathcal{R}_{t+1}^{(r)}$ and $\{V^{k_1}, \dots, V^{k_{t+1}}\}$.*

Proof of Claim 5. To begin with, consider $t = d_r$. As x is Pareto-optimal with respect to \mathcal{S} and $\{V^1, \dots, V^{d+1}\}$, x is also Pareto-optimal with respect to $\bigcap_{k \in [d] \setminus K} \mathcal{S}_{P_k}(x)$ and

$$\{V^k : k \in K \cup (d+1)\} = \{V^{k_1}, \dots, V^{k_{d_r+1}}\}$$

due to Lemma 10. Consequently, x is also Pareto-optimal with respect to $\mathcal{R}_{d_r+1}^{(r)} = \mathcal{S}_I(x) \cap \bigcap_{k \in [d] \setminus K} \mathcal{S}_{P_k}(x)$ and $\{V^{k_1}, \dots, V^{k_{d_r+1}}\}$ due to Proposition 7. Note that this property is even stronger than weak Pareto-optimality. We will need this strong version in the induction step when $t = d_r - 1$.

Now consider a iteration $t \leq d_r - 1$ that is reached and assume that the induction hypothesis is true for $t + 1$. We consider iteration $t + 1$ where $\mathcal{R}_{t+1}^{(r)}$ is defined, and distinguish between two cases. If $\mathcal{C}_{t+1}^{(r)} = \emptyset$, then x is weakly Pareto-optimal with respect to $\mathcal{R}_{t+1}^{(r)} = \mathcal{R}_{t+2}^{(r)} \cap \mathcal{S}_I(x)$ and $\{V^{k_1}, \dots, V^{k_{t+1}}\}$ due to the induction hypothesis, Lemma 8 (I), and Proposition 7.

Let us consider the more interesting case $\mathcal{C}_{t+1}^{(r)} \neq \emptyset$. Since iteration t is reached, there is no call of the Witness_0 function in iteration $t + 1$, i.e., $K_{\text{eq}} = ()$ in iteration $t + 1$. The induction hypothesis and Lemma 8 (II) yield $V^{k_{t+2}}x \leq V^{k_{t+2}}x^{(r, t+1)}$.

We will show that even $V^{k_{t+2}}x < V^{k_{t+2}}x^{(r, t+1)}$. For this, we assume to the contrary that $V^{k_{t+2}}x = V^{k_{t+2}}x^{(r, t+1)}$ and distinguish between the cases $t = d_r - 1$ and $t < d_r - 1$.

1. If $t = d_r - 1$, then we obtain $V^{k_{d_r+1}}x = V^{k_{d_r+1}}x^{(r, d_r)}$ and $V^{k_1 \dots k_{d_r}}x^{(r, d_r)} < V^{k_1 \dots k_{d_r}}x$ since $x^{(r, d_r)} \in \mathcal{C}_{d_r}^{(r)}$. Hence, $x^{(r, d_r)} \in \mathcal{R}_{d_r+1}^{(r)}$ dominates x with respect to $\{V^{k_1}, \dots, V^{k_{d_r+1}}\}$ which contradicts the fact that x is Pareto-optimal with respect to $\mathcal{R}_{d_r+1}^{(r)}$ and $\{V^{k_1}, \dots, V^{k_{d_r+1}}\}$.
2. If $t < d_r - 1$, then $V^{k_{t+2}}x = V^{k_{t+2}}x^{(r, t+1)}$ implies $x|_{P_{k_{t+2}}} = x^{(r, t)}|_{P_{k_{t+2}}}$ as we assume that the OK_0 -event occurs. Consequently, $k_{t+2} \in K_{\text{eq}}$ in iteration $t + 1$, which contradicts the previous observation that $K_{\text{eq}} = ()$ in that iteration.

This concludes the proof of Claim 5. □

With Claim 5 we are now able to show that a call $\text{Witness}_0(V, x, K, r, I)$ terminates without a further call to the Witness_0 function if and only if $K = ()$. Note that one direction is trivial.

Claim 6. *Consider an arbitrary call $\text{Witness}_0(V, x, K, r, I)$. If $K \neq ()$, then this call results in another call to the Witness_0 function (and does not terminate in Line 34).*

Proof of Claim 6. Let us assume that there is no further call to the Witness_0 function until iteration $t = 0$, i.e., we reach iteration $t = 0$. In accordance with Claim 5, x is weakly Pareto-optimal with respect to $\mathcal{R}_1^{(r)}$ and $\{V^{k_1}\}$. Now let us consider iteration $t = 0$. We obtain $\mathcal{C}_0^{(r)} = \mathcal{R}_1^{(r)}$ since there are no restrictions in this iteration. Consequently, $\mathcal{C}_0^{(r)} \neq \emptyset$ because $x \in \mathcal{R}_1^{(r)}$. The solution $x^{(r, 0)}$ minimizes V^{k_1} among all solutions from $\mathcal{C}_0^{(r)}$. On the other hand, $x^{(r, 0)}$ cannot dominate x strongly. Hence, $V^{k_1}x^{(r, 0)} = V^{k_1}x$, i.e., $x^{(r, 0)}|_{P_{k_1}} = x|_{P_{k_1}}$ as we assumed that the OK_0 -event occurs. Therefore, $k_1 \in K_{\text{eq}}$, i.e., $K_{\text{eq}} \neq ()$, and thus, the Witness_0 function is called in Line 23. □

According to Claim 6 there will be recursive calls until a call of the form $\text{Witness}_0(V, x, (), r, I)$. This call immediately returns the set $\mathcal{R}_{d_{r+1}}^{(r)}$ in Line 4. Since $[d] \setminus () = [d]$, we obtain

$$\mathcal{R}_{d_{r+1}}^{(r)} = \mathcal{S}_I(x) \cap \bigcap_{k \in [d]} \mathcal{S}_{P_k}(x) = \mathcal{S}_{[n]}(x) = \{x\}.$$

Now consider call number $r - 1$ and the iteration t_{r-1} in this call in which $\text{Witness}_0(V, x, (), r, I)$ has been called. In this iteration, $K_{\text{eq}} \neq ()$ since Line 23 is reached. Hence, there is at least one index $k \in K_{\text{eq}}$, and for these indices, r_k is set to r in Line 13. Now, as the next call is of the form $\text{Witness}_0(V, x, (), r, I)$, this implies $K_{\text{neq}} = ()$, i.e., all values r_k for $k \in [d]$ have been set by now and, thus,

$$r^* = \max\{r_1, \dots, r_d\} = r - 1,$$

i.e., the number of the call we currently consider.

Consider the solution $x^{(r^*, t_{r^*})}$ defined in Line 8 and let K be the tuple of call r^* . Since $K_{\text{neq}} = ()$ this implies $K_{\text{eq}} = K$. Hence, $x^{(r^*, t_{r^*})}|_{P_k} = x|_{P_k}$ for all $k \in K$ by definition of K_{eq} in Line 9. On the other hand,

$$x^{(r^*, t_{r^*})} \in \mathcal{C}_{t_{r^*}}^{(r^*)} \subseteq \mathcal{R}_{t_{r^*}+1}^{(r^*)} \subseteq \mathcal{R}_{d_{r^*}+1}^{(r^*)} \subseteq \bigcap_{k \in [d] \setminus K} \mathcal{S}_{P_k}(x).$$

The first inclusion is due to the definition of $\mathcal{C}_{t_{r^*}}^{(r^*)}$ in Line 6. The second inclusion is due to the observation that always $\mathcal{R}_{t_{i+1}}^{(r)} \subseteq \mathcal{R}_{t_{i+2}}^{(r)} \subseteq \dots \subseteq \mathcal{R}_{d_{r+1}}^{(r)}$ due to the construction of the sets $\mathcal{R}_t^{(r)}$ in Line 20 and Line 31. The construction of $\mathcal{R}_{d_{r^*}+1}^{(r^*)}$ in Line 3 yields the third inclusion. Hence, $x^{(r^*, t_{r^*})}|_{P_k} = x|_{P_k}$ for all $k \in [d] \setminus K$, and according to the previous observations, even for all $k \in K$. Consequently, $x^{(r^*, t_{r^*})} = x$. Summarizing the previous results, we obtain that the call $\text{Witness}_0(V, x, [d], 1, ())$ ends up in the call $\text{Witness}_0(V, x, (), r, I)$ for some index tuple I which immediately returns the set $\mathcal{R}_{d_{r+1}}^{(r)} = \{x\} = \{x^{(r^*, t_{r^*})}\}$. \square \square

Like for the simple Witness function, we show that it is enough to know some information about the run of the Witness_0 function to reconstruct the solution x . As before, we call this data the *certificate* of x .

Definition 31. Let r_1, \dots, r_d and t_1, \dots, t_{r^*} for $r^* = \max\{r_1, \dots, r_d\}$ be the indices and $x^{(r, t)}$ be the vectors constructed during the execution of $\text{Witness}_0(V, x, [d], 1, ())$. Furthermore, consider the tuple I at the moment when the last call to the Witness_0 function terminates. The pair (I^*, A) for $I^* = I \cup (i_1^*, \dots, i_d^*)$, $i_k^* = \min(P_k \setminus I)$, and $A = [x^{(1, d_1)}, \dots, x^{(1, t_1)}, \dots, x^{(r^*, d_{r^*})}, \dots, x^{(r^*, t_{r^*})}]|_{I^*}$, is called the *V-certificate* of x . We label the columns of A by $a^{(r, t)}$. Moreover, we call a pair (I', A') a *certificate* if there is some realization V such that the OK_0 -events occurs and if there exists a Pareto-optimal solution $x \in \mathcal{S}$ such that (I', A') is the *V-certificate* of x . By \mathcal{C} we denote the set of all certificates.

We assume that the indices r_k and t_r (and hence also the indices d_r) are implicitly encoded in a given certificate. Later we will take these indices into consideration again when we count the number of possible certificates.

Lemma 32. Let V be a realization for which the OK_0 -event occurs and let (I^*, A) be a *V-certificate* of some Pareto-optimal solution x . Let A be of the form

$$A = [a^{(1, d_1)}, \dots, a^{(1, t_1)}, \dots, a^{(r^*, d_{r^*})}, \dots, a^{(r^*, t_{r^*})}].$$

For a fixed index $k \in [d]$ let

$$M = [a^{(1, d_1)}, \dots, a^{(1, t_1)}, \dots, a^{(r_k, d_{r_k})}, \dots, a^{(r_k, t_{r_k})}]|_J,$$

where $J = I^* \cap P_k = (j_1, \dots, j_m)$. Then M is of the form

$$M = \begin{bmatrix} \bar{x}_{j_1} & x_{j_1} & \dots & x_{j_1} \\ * & \ddots & \ddots & \vdots \\ \vdots & \ddots & \bar{x}_{j_{m-1}} & x_{j_{m-1}} \\ * & \dots & * & x_{j_m} \end{bmatrix} \in \{0, \dots, \mathcal{K}\}^{|J| \times |J|},$$

where each ‘*’ can be an arbitrary value from $\{0, \dots, \mathcal{K}\}$ (different ‘*’-entries can represent different values) and where \bar{z} for a value $z \in \{0, \dots, \mathcal{K}\}$ can be an arbitrary value from $\{0, \dots, \mathcal{K}\} \setminus \{z\}$.

Proof. Consider the call $\text{Witness}_0(V, x, [d], 1, ())$ and all subsequent calls $\text{Witness}_0(V, x, K, r, I)$. By definition of r_k we have $r \leq r_k \iff k \in K$ (see Line 13, Line 10, and Line 23). In each call where $r \leq r_k$ one vector $x^{(r,t)}$ is constructed in each iteration t . Also, in each iteration except for the last iteration t_{r_k} of the r_k^{th} call, when $k \in K_{\text{eq}}$ for the first and the last time, one index $i \in P_k$ is chosen and added to I . Since J consists of the chosen indices $i \in P_k$ and the additional index i_k^* , matrix M is a square matrix.

We first consider the last column of M . As $x^{(r_k, t_{r_k})}$ is the last vector constructed before k is removed from K , index k must be an element of K_{eq} in iteration t_{r_k} of call r_k , i.e., $x^{(r_k, t_{r_k})}|_{P_k} = x|_{P_k}$. Hence, the last column of M has the claimed form because $J \subseteq P_k$.

Now consider the remaining columns of M . Due to the construction of the set $\mathcal{R}_t^{(r)}$ in Line 3, Line 20, and Line 31, all vectors $z \in \mathcal{R}_t^{(r)}$ coincide with x in the previously chosen indices i . As in the case $\mathcal{C}_t^{(r)} \neq \emptyset$ vector $x^{(r,t)}$ is an element of $\mathcal{C}_t^{(r)} \subseteq \mathcal{R}_{t+1}^{(r)}$ and in the case $\mathcal{C}_t^{(r)} = \emptyset$ vector $x^{(r,t)}$ is constructed appropriately, the upper triangle of M , excluding the principal diagonal, has the claimed form. The form of the principal diagonal follows from the choice of index $i \in P_k$: In Line 15 we chose i such that $x_i^{(r,t)} \neq x_i$, in Line 30 we construct $x^{(r,t)}$ explicitly such that $x_i^{(r,t)} \neq x_i$. \square

Like in the model without zero-preserving perturbations, our goal is to execute the Witness_0 function without revealing the entire matrix V . For this we now consider a variant of the Witness_0 function given as Algorithm 4 which gets as additional parameters the V -certificate of x , a shift vector $u \in \{0, \dots, \mathcal{K}\}^n$, the ε -box $B = B_V(x - u)$, and a set \mathcal{S}' of solutions that are still under consideration. Recall that at the beginning of every call to the original Witness_0 function the set of solutions that have still to be considered is restricted to a subset of $\bigcap_{k \in [d] \setminus K} \mathcal{S}_{P_k}(x)$ (see Line 3). The huge amount of information that is necessary to restrict the current set of solutions to those that still have to be considered is not given by the V -certificate of x . Thus, we keep track of this set of remaining solutions by passing it as a parameter. We will see how to update this set without too much knowledge about x (cf. Line 13 of Algorithm 4).

It is important to break ties in Line 9 the same way as we did in the original Witness_0 function, i.e., we take the lexicographically first solution.

Lemma 33. *Let (I^*, A) be the V -certificate of x , let $u \in \{0, \dots, \mathcal{K}\}^n$ be an arbitrary vector, and let $B = B_V(x - u)$. Then the call $\text{Witness}_0(V, [d], 1, I^*, A, \mathcal{S}, B, u)$ returns $\{x\}$.*

Before we give a formal proof of Lemma 33 we try to give some intuition for it. As for the simple variant of the Witness function we restrict the set of solutions to vectors that look like the vectors we want to reconstruct in the next iterations of the current call, i.e., we intersect the current set with the set $\bigcup_{s=t_r}^{t-1} \mathcal{S}_{I^*}(a^{(r,s)})$ in iteration t . In this way we only deal with subsets of the original sets, but we do not lose the vectors we want to reconstruct. In order to reconstruct the vectors, we need more information than in the simple variant: we need to know in which iterations the recursive calls of Witness_0 are made, in each call we need to know which objective functions V^k must not be considered anymore, and for each of these objective functions we need to know the vector $x|_{P_k}$. The information when the recursive calls are made and which objective functions

Algorithm 4: $\text{Witness}_0(V, K, r, I^*, A, \mathcal{S}', B, u)$

```
1 let  $d_r$  be the number of components of  $K$  and let  $K$  be of the form  $K = (k_1, \dots, k_{d_r})$  ;
2 set  $k_{d_r+1} = d + 1$  ;
3 let  $b$  be the corner of  $B$  ;
4 if  $d_r = 0$  then return  $\mathcal{S}'$  ;
5 set  $\mathcal{R}_{d_r+1}^{(r)} = \mathcal{S}' \cap \bigcup_{s=t_r}^{d_r} \mathcal{S}_{I^*}^{(r,s)}(a^{(r,s)})$  ;
6 for  $t = d_r, d_r - 1, \dots, 0$  do
7   set  $\mathcal{C}_t^{(r)} = \{z \in \mathcal{R}_{t+1}^{(r)} : V^{k_1 \dots k_t} \cdot (z - u) \leq b|_{k_1 \dots k_t}\} \cap \mathcal{S}_{I^*}^{(r,t)}(a^{(r,t)})$  ;
8   if  $\mathcal{C}_t^{(r)} \neq \emptyset$  then
9     set  $x^{(r,t)} = \arg \min \{V^{k_{t+1}} z : z \in \mathcal{C}_t^{(r)}\}$  ;
10    if  $t = t_r$  then
11      let  $K_{\text{eq}} \subseteq K$  be the tuple of indices  $k$  for which  $r_k = r$  ;
12      set  $K_{\text{neq}} = K \setminus K_{\text{eq}}$  ;
13      return  $\text{Witness}_0(V, K_{\text{neq}}, r + 1, I^*, A, \mathcal{S}' \cap \bigcap_{k \in K_{\text{eq}}} \mathcal{S}_{P_k}(x^{(r,t)}), B, u)$  ;
14    else
15      set  $\mathcal{R}_t^{(r)} = \{z \in \mathcal{R}_{t+1}^{(r)} : V^{k_{t+1}} z < V^{k_{t+1}} x^{(r,t)}\} \cap \bigcup_{s=t_r}^{t-1} \mathcal{S}_{I^*}^{(r,s)}(a^{(r,s)})$  ;
16    end
17  else
18    set  $x^{(r,t)} = (\perp, \dots, \perp)$  ;
19    set  $\mathcal{R}_t^{(r)} = \mathcal{R}_{t+1}^{(r)} \cap \bigcup_{s=t_r}^{t-1} \mathcal{S}_{I^*}^{(r,s)}(a^{(r,s)})$  ;
20  end
21 end
22 return  $\emptyset$  ;
```

must not be considered anymore is given in the certificate: The variable t_r contains the iteration number when the recursive call is made. The index r_k contains the number of the call where index k has to be removed from K . Hence, index k is removed in the $t_{r_k}^{\text{th}}$ iteration of call r_k . If we can reconstruct K_{eq} and the vector $x^{(r,t)}$ in the iteration where we make the recursive call, then we can also reconstruct the bits of x at indices $i \in P_k$ for all indices $k \in K_{\text{eq}}$ because $x|_{P_k} = x^{(r,t)}|_{P_k}$ for these indices k (cf. Line 13).

Proof. We compare the executions of $\text{Witness}_0(V, x, [d], 1, ())$ and $\text{Witness}_0(V, [d], 1, I^*, A, \mathcal{S}, B, u)$ and show the following claim by induction on r .

Claim 7. *If there is a call of the form $\text{Witness}_0(V, x, K, r, I)$ during the execution of the call $\text{Witness}_0(V, x, [d], 1, ())$, then there is also a call of the form $\text{Witness}_0(V, K, r, I^*, A, \mathcal{S}', B, u)$ for $\mathcal{S}' = \bigcap_{k \in [d] \setminus K} \mathcal{S}_{P_k}(x)$ during the execution of the call $\text{Witness}_0(V, [d], 1, I^*, A, \mathcal{S}, B, u)$.*

Proof of Claim 7. For the case $r = 1$ it is true if we recall the convention that $\bigcap_{k \in ()} \mathcal{S}_{P_k}(x) = \mathcal{S}$. Now let us consider an arbitrary call $r + 1$ and assume that Claim 7 holds for r . Hence, we can assume that there are calls of the form $\text{Witness}_0(V, x, K, r, I)$ and $\text{Witness}_0(V, K, r, I^*, A, \mathcal{S}', B, u)$ for $\mathcal{S}' = \bigcap_{k \in [d] \setminus K} \mathcal{S}_{P_k}(x)$. We now show that both calls are executed essentially the same way. Formally, we prove the following claims by induction on t , where $\mathcal{R}_t^{(r)}$, $\mathcal{C}_t^{(r)}$, $x^{(r,t)}$, K'_{eq} , and K'_{neq} refer to the sets, vectors, and tuples from the call $\text{Witness}_0(V, K, r, I^*, A, \mathcal{S}', B, u)$.

Claim 8. $\mathcal{R}_t^{(r)} \subseteq \mathcal{R}_t^{(r)}$ for all $t \in \{t_r + 1, \dots, d_r + 1\}$.

Claim 9. $x^{(r,t)} = x^{(r,t)}$ for all $t \in \{t_r, \dots, d_r\}$ for which $\mathcal{C}_t^{(r)} \neq \emptyset$.

Claim 10. $x^{(r,s)} \in \mathcal{R}'_t^{(r)}$ for all $t \in \{t_r + 1, \dots, d_r + 1\}$ and all $s \in \{t_r, \dots, t - 1\}$ for which $\mathcal{C}_s^{(r)} \neq \emptyset$.

Proof of Claim 8, Claim 9, and Claim 10. We apply a downward induction on t . For the beginning, consider $t = d_r + 1$. We have

$$\begin{aligned}\mathcal{R}'_{d_r+1} &= \mathcal{S}' \cap \bigcup_{s=t_r}^{d_r} \mathcal{S}_{I^*}(a^{(r,s)}) \text{ for } \mathcal{S}' = \bigcap_{k \in [d] \setminus K} \mathcal{S}_{P_k}(x) \quad \text{and} \\ \mathcal{R}_{d_r+1} &= \mathcal{S}_I(x) \cap \bigcap_{k \in [d] \setminus K} \mathcal{S}_{P_k}(x).\end{aligned}$$

Due to the construction of the vectors $x^{(r,s)}$ and the definition of $a^{(r,s)}$,

$$a^{(r,s)}|_I = x^{(r,s)}|_I = x|_I$$

for all $s = t_r, \dots, d_r$ (see Lemma 32). The inclusion $I^* \supseteq I$ yields

$$\mathcal{S}_{I^*}(a^{(r,s)}) \subseteq \mathcal{S}_I(a^{(r,s)}) = \mathcal{S}_I(x)$$

for all $s = t_r, \dots, d_r$. Consequently, $\mathcal{R}'_{d_r+1} \subseteq \mathcal{R}_{d_r+1}^{(r)}$. For Claim 9 nothing has to be shown in the initial step $t = d_r + 1$ of the induction. Let us consider Claim 10 and let $s \in \{t_r, \dots, d_r\}$ be an arbitrary index for which $\mathcal{C}_s^{(r)} \neq \emptyset$. Then

$$x^{(r,s)} \in \mathcal{C}_s^{(r)} \subseteq \mathcal{R}_{s+1}^{(r)} \subseteq \mathcal{R}_{d_r+1}^{(r)} \subseteq \bigcap_{k \in [d] \setminus K} \mathcal{S}_{P_k}(x).$$

Furthermore, $x^{(r,s)} \in \mathcal{S}_J(a^{(r,s)})$ for every index tuple J due to the definition of $a^{(r,s)}$. Consequently, $x^{(r,s)} \in \mathcal{R}'_{d_r+1}^{(r)}$.

For the induction step let $t \leq d_r$. Due to the occurrence of the OK_0 -event and the fact that $B = B_V(x - u)$, we obtain

$$\begin{aligned}\mathcal{C}'_t &= \{z \in \mathcal{R}'_{t+1}^{(r)} : V^{k_1 \dots k_t} \cdot (z - u) \leq b|_{k_1 \dots k_t}\} \cap \mathcal{S}_{I^*}(a^{(r,t)}) \\ &= \{z \in \mathcal{R}'_{t+1}^{(r)} : V^{k_1 \dots k_t} z < V^{k_1 \dots k_t} x\} \cap \mathcal{S}_{I^*}(a^{(r,t)}) \quad \text{and} \\ \mathcal{C}_t &= \{z \in \mathcal{R}_{t+1}^{(r)} : V^{k_1 \dots k_t} z < V^{k_1 \dots k_t} x\}.\end{aligned}$$

Since $\mathcal{R}'_{t+1}^{(r)} \subseteq \mathcal{R}_{t+1}^{(r)}$, we obtain $\mathcal{C}'_t \subseteq \mathcal{C}_t$. First, we consider the case $\mathcal{C}'_t = \emptyset$ which implies $\mathcal{C}_t = \emptyset$ and $t \geq t_r + 1$. The inequality follows from the fact that in iteration t_r the Witness_0 function is called (Line 23 of Algorithm 3) which implies $\mathcal{C}'_{t_r} \neq \emptyset$. In this case,

$$\begin{aligned}\mathcal{R}'_t &= \mathcal{R}'_{t+1}^{(r)} \cap \bigcup_{s=t_r}^{t-1} \mathcal{S}_{I^*}(a^{(r,s)}) \quad \text{and} \\ \mathcal{R}_t &= \mathcal{R}_{t+1}^{(r)} \cap \mathcal{S}_I(x),\end{aligned}$$

where I is the updated index tuple I . Due to the construction of the vectors $x^{(r,s)}$ (see Lemma 32) and the definition of the vectors $a^{(r,s)}$, we know that

$$a^{(r,s)}|_I = x^{(r,s)}|_I = x|_I$$

for all $s = t_r, \dots, t - 1$. As $I^* \supseteq I$, this implies

$$\bigcup_{s=t_r}^{t-1} \mathcal{S}_{I^*}(a^{(r,s)}) \subseteq \bigcup_{s=t_r}^{t-1} \mathcal{S}_I(a^{(r,s)}) = \mathcal{S}_I(x).$$

As $\mathcal{R}'_{t+1} \subseteq \mathcal{R}_{t+1}^{(r)}$ in accordance with the induction hypothesis, Claim 8, we obtain $\mathcal{R}'_t \subseteq \mathcal{R}_t^{(r)}$. For Claim 9 nothing has to be shown here. Let $s \in \{t_r, \dots, t-1\}$ be an arbitrary index for which $\mathcal{C}_s^{(r)} \neq \emptyset$. Then $x^{(r,s)} \in \mathcal{R}'_{t+1}^{(r)}$ by Claim 10 of the induction hypothesis, $x^{(r,s)} \in \mathcal{S}_{I^*}(a^{(r,s)})$, and consequently $x^{(r,s)} \in \mathcal{R}'_t$.

Let us finally consider the case $\mathcal{C}_t^{(r)} \neq \emptyset$. Claim 10 of the induction hypothesis yields $x^{(r,t)} \in \mathcal{R}'_{t+1}^{(r)}$. Since $x^{(r,t)} \in \mathcal{S}_{I^*}(a^{(r,t)})$ and $V^{k_1 \dots k_t} x^{(r,t)} < V^{k_1 \dots k_t} x$, also $x^{(r,t)} \in \mathcal{C}'_t$ and, thus, $\mathcal{C}'_t \neq \emptyset$. Hence, $x^{(r,t)} = x^{(r,t)}$ as $\mathcal{C}'_t \subseteq \mathcal{C}_t^{(r)}$. Claim 8 and Claim 10 have only to be validated if $t \geq t_r$, i.e., we can assume that $K_{\text{eq}} = ()$. Then

$$\mathcal{R}'_t = \{z \in \mathcal{R}'_{t+1} : V^{k_{t+1}} z < V^{k_{t+1}} x^{(r,t)}\} \cap \bigcup_{s=t_r}^{t-1} \mathcal{S}_{I^*}(a^{(r,s)})$$

because $x^{(r,t)} = x^{(r,t)}$, and

$$\mathcal{R}_t^{(r)} = \{z \in \mathcal{R}_{t+1}^{(r)} : V^{k_{t+1}} z < V^{k_{t+1}} x^{(r,t)}\} \cap \mathcal{S}_I(x).$$

With the same argument used for the case $\mathcal{C}_t^{(r)} = \emptyset$ we obtain

$$\mathcal{R}'_{t+1} \cap \bigcup_{s=t_r}^{t-1} \mathcal{S}_{I^*}(a^{(r,s)}) \subseteq \mathcal{R}_{t+1}^{(r)} \cap \mathcal{S}_I(x)$$

and, hence, $\mathcal{R}'_t \subseteq \mathcal{R}_t^{(r)}$. Consider an arbitrary index $s \in \{t_r, \dots, t-1\}$ for which $\mathcal{C}_s^{(r)} \neq \emptyset$. Then

$$x^{(r,s)} \in \mathcal{C}_s^{(r)} \subseteq \mathcal{R}_{s+1}^{(r)} \subseteq \mathcal{R}'_t.$$

In particular, $V^{k_{t+1}} x^{(r,s)} < V^{k_{t+1}} x^{(r,t)}$ (see Line 20) and, hence, $V^{k_{t+1}} x^{(r,s)} < V^{k_{t+1}} x^{(r,t)}$ because $x^{(r,t)} = x^{(r,t)}$. Furthermore, $x^{(r,s)} \in \mathcal{R}'_{t+1}^{(r)}$ due to the induction hypothesis, Claim 10, and $x^{(r,s)} \in \mathcal{S}_{I^*}(a^{(r,s)})$. Consequently, $x^{(r,s)} \in \mathcal{R}'_t$. \square

The induction step of the proof of Claim 7 follows from the three claims above: Let us assume that there is a call of the form $\text{Witness}_0(V, x, \hat{K}, r+1, \hat{I})$. By the definition of t_r , this call is executed in iteration t_r of call r . Consequently, $x^{(r,t_r)} \in \mathcal{C}'_{t_r} \neq \emptyset$. Applying Claim 10 for $s = t_r$ and $t = t_r + 1$, we obtain $x^{(r,t_r)} \in \mathcal{R}'_{t_r+1}^{(r)}$. As $x^{(r,t_r)} \in \mathcal{C}'_{t_r}$, the inequalities $V^{k_1 \dots k_{t_r}} x^{(r,t_r)} < V^{k_1 \dots k_{t_r}} x$ hold, which are equivalent to $V^{k_1 \dots k_{t_r}} \cdot (x^{(r,t_r)} - u) \leq b|_{k_1 \dots k_{t_r}}$ due to the occurrence of the OK₀-event. Furthermore, $x^{(r,t_r)} \in \mathcal{S}_{I^*}(a^{(r,t_r)})$ by the definition of $a^{(r,t_r)}$. Hence, $x^{(r,t_r)} \in \mathcal{C}'_{t_r}$ (see Line 7), i.e., $\mathcal{C}'_{t_r} \neq \emptyset$. Moreover, $x^{(r,t_r)} = x^{(r,t_r)}$ in accordance with Claim 9. In iteration t_r of the call $\text{Witness}_0(V, K, r, I^*, A, \mathcal{S}', B, u)$ Line 11 is reached. By the definition of the values r_k we obtain $K'_{\text{eq}} = K_{\text{eq}}$, and hence, $x^{(r,t_r)}|_{P_k} = x^{(r,t_r)}|_{P_k} = x|_{P_k}$ for all $k \in K'_{\text{eq}} = K_{\text{eq}}$ due to the definition of K_{eq} . In Line 13, there is a call of the form $\text{Witness}_0(V, K'_{\text{neq}}, r+1, I^*, A, \mathcal{S}' \cap \bigcap_{k \in K'_{\text{eq}}} \mathcal{S}_{P_k}(x^{(r,t_r)}), B, u)$. The correctness of Claim 7 follows because

$$K'_{\text{neq}} = K \setminus K'_{\text{eq}} = K \setminus K_{\text{eq}} = K_{\text{neq}} = \hat{K},$$

where \hat{K} is the parameter from the call $\text{Witness}_0(V, x, \hat{K}, r+1, \hat{I})$, and because

$$\begin{aligned} \mathcal{S}' \cap \bigcap_{k \in K'_{\text{eq}}} \mathcal{S}_{P_k}(x^{(r,t_r)}), B, u &= \bigcap_{k \in [d] \setminus K} \mathcal{S}_{P_k}(x) \cap \bigcap_{k \in K'_{\text{eq}}} \mathcal{S}_{P_k}(x^{(r,t_r)}) \\ &= \bigcap_{k \in [d] \setminus K} \mathcal{S}_{P_k}(x) \cap \bigcap_{k \in K'_{\text{eq}}} \mathcal{S}_{P_k}(x) \end{aligned}$$

$$= \bigcap_{k \in [d] \setminus K'_{\text{neq}}} \mathcal{S}_{P_k}(x). \quad \square$$

□

Let us finish the proof of Lemma 33. According to Lemma 30, the call $\text{Witness}_0(V, x, [d], 1, ())$ returns the set $\{x\} \neq \emptyset$. Hence, there is a call of the form $\text{Witness}_0(V, x, K, r, I)$ for $K = ()$ (see Line 4). Due to Claim 7, there must be also a call of the form $\text{Witness}_0(V, (), r, I^*, A, \mathcal{S}', B, u)$ for $\mathcal{S}' = \bigcap_{k \in [d] \setminus ()} \mathcal{S}_{P_k}(x) = \{x\}$. This set is immediately returned in Line 4. □ □

By the choice of the vector u we can control which information about V has to be known in order to be able to execute the call $\text{Witness}_0(V, [d], 1, I^*, A, \mathcal{S}, B, u)$. While Lemma 33 is correct for every choice of $u \in \{0, \dots, \mathcal{K}\}^n$, we have to choose u carefully in order for the following probabilistic analysis to work. Later we will see that $u^* = u^*(I^*, A)$, given by

$$u_i^* = \begin{cases} |x_i - 1| & \text{if } i \in (i_1^*, \dots, i_d^*), \\ x_i & \text{if } i \in I^* \setminus (i_1^*, \dots, i_d^*), \\ 0 & \text{otherwise,} \end{cases} \quad (3)$$

is well-suited for our purpose. Recall that $i_k^* \in P_k$ are the indices that have been added to I in the definition of the V -certificate to obtain I^* . Furthermore, x_i is given by the last column of A for every index $i \in I^*$ (cf. Lemma 30). Hence, vector u^* can be defined with the information that is contained in the V -certificate of x ; we do not have to know the vector x itself.

In the next step, we bound the number of Pareto-optimal solutions. For this, consider the following function $\chi_{I^*, A, B}(V)$, parameterized by an arbitrary certificate $(I^*, A) \in \mathcal{C}$ and an arbitrary ε -box $B \in \mathbb{B}_\varepsilon$, that is defined as follows: $\chi_{I^*, A, B}(V) = 1$ if $\text{Witness}_0(V, [d], 1, I^*, A, \mathcal{S}, B, u^*(I^*, A))$ returns a set $\{x'\}$ such that $B_V(x' - u^*(I^*, A)) = B$, and $\chi_{I^*, A, B}(V) = 0$ otherwise.

Corollary 34. *Assume that the OK_0 -event occurs. Then the number $PO(V)$ of Pareto-optimal solutions is at most*

$$\sum_{(I^*, A) \in \mathcal{C}} \sum_{B \in \mathbb{B}_\varepsilon} \chi_{I^*, A, B}(V).$$

Proof. Let x be a Pareto-optimal solution, let (I^*, A) be the V -certificate of x , and let $B = B_V(x - u^*(I^*, A)) \in \mathbb{B}_\varepsilon$. Due to Lemma 33, $\text{Witness}_0(V, [d], 1, I^*, A, \mathcal{S}, B, u^*(I^*, A))$ returns $\{x\}$. Hence, $\chi_{I^*, A, B}(V) = 1$. It remains to show that this function $x \mapsto (I^*, A, B')$ defined within the previous lines is injective. Let x_1 and x_2 be distinct Pareto-optimal solutions and let (I_1^*, A_1) and (I_2^*, A_2) be the V -certificates of x_1 and x_2 , respectively. If $(I_1^*, A_1) \neq (I_2^*, A_2)$, then x_1 and x_2 are mapped to distinct triplets. Otherwise, $u^*(I_1^*, A_1) = u^*(I_2^*, A_2)$ and, hence, $B_V(x_1 - u^*(I_1^*, A_1)) \neq B_V(x_2 - u^*(I_2^*, A_2))$ because of the OK_0 -event. Consequently, also in this case x_1 and x_2 are mapped to distinct triplets. □

Corollary 34 immediately implies a bound on the expected number of Pareto-optimal solutions.

Corollary 35. *The expected number of Pareto-optimal solutions is bounded by*

$$\mathbf{E}_V[PO(V)] \leq \sum_{(I^*, A) \in \mathcal{C}} \sum_{B \in \mathbb{B}_\varepsilon} \Pr_V[E_{I^*, A, B}] + (\mathcal{K} + 1)^n \cdot \Pr_V[\overline{OK_0(V)}]$$

where $E_{I^*, A, B}$ denotes the event that the call $\text{Witness}_0(V, [d], 1, I^*, A, \mathcal{S}, B, u^*(I^*, A))$ returns a set $\{x'\}$ such that $B_V(x' - u^*(I^*, A)) = B$.

Proof. By applying Corollary 34, we obtain

$$\begin{aligned}
& \mathbf{E}_V[\text{PO}(V)] \\
&= \mathbf{E}_V[\text{PO}(V) \mid \text{OK}_0(V)] \cdot \mathbf{Pr}_V[\text{OK}_0(V)] + \mathbf{E}_V[\text{PO}(V) \mid \overline{\text{OK}_0(V)}] \cdot \mathbf{Pr}_V[\overline{\text{OK}_0(V)}] \\
&\leq \mathbf{E}_V \left[\sum_{(I^*, A) \in \mathcal{C}} \sum_{B \in \mathbb{B}_\varepsilon} \chi_{I^*, A, B}(V) \mid \text{OK}_0(V) \right] \cdot \mathbf{Pr}_V[\text{OK}_0(V)] + |\mathcal{S}| \cdot \mathbf{Pr}_V[\overline{\text{OK}_0(V)}] \\
&\leq \mathbf{E}_V \left[\sum_{(I^*, A) \in \mathcal{C}} \sum_{B \in \mathbb{B}_\varepsilon} \chi_{I^*, A, B}(V) \right] + (\mathcal{K} + 1)^n \cdot \mathbf{Pr}_V[\overline{\text{OK}_0(V)}] \\
&= \sum_{(I^*, A) \in \mathcal{C}} \sum_{B \in \mathbb{B}_\varepsilon} \mathbf{Pr}_V[E_{I^*, A, B}] + (\mathcal{K} + 1)^n \cdot \mathbf{Pr}_V[\overline{\text{OK}_0(V)}]. \quad \square
\end{aligned}$$

We will see that the first term of the sum in Corollary 35 can be bounded independently of ε and that the second term tends to 0 for $\varepsilon \rightarrow 0$. First of all, we analyze the size of the certificate space.

Lemma 36. *The size of the certificate space is bounded by*

$$|\mathcal{C}| = (\mathcal{K} + 1)^{(d^2+d)(d^3+d^2+d)} \cdot O(n^{d^3+d^2}).$$

Proof. Consider the execution of the call $\text{Witness}_0(V, x, [d], 1, ())$ and let $r^* = \max\{r_1, \dots, r_d\}$ be the maximum of the values r_1, \dots, r_d . Including this call with number 1, there can be at most d calls to the Witness_0 function except for the call with number $r^* + 1$ that terminates due to $d_{r^*+1} = 0$. This is because in each of the other calls at least one index $k \in [d]$ is removed from the tuple K . Hence, $r_1, \dots, r_d \in [d]$. Consequently, there are at most d^d possibilities for these numbers. In the r^{th} call, the iteration number t_r is an element of $[d_r]_0 \subseteq [d]_0$, and hence, there are at most $(d+1)^{r^*} \leq (d+1)^d$ possibilities to choose iteration numbers t_1, \dots, t_{r^*} . In each iteration, at most d indices i are added to the tuple I . As there are at most d calls and at most $d+1$ iterations each call, tuple I contains at most $d^2 \cdot (d+1)$ indices in total. Hence, there are at most

$$\sum_{k=1}^{d^2(d+1)} n^k \leq d^2 \cdot (d+1) \cdot n^{d^2 \cdot (d+1)}$$

choices for I . Once I is fixed, also the indices in $I^* \setminus I$ are fixed because the indices added to I in Definition 31 are determined by I . The tuple I^* contains $|I| + d \leq d^3 + d^2 + d$ indices. In each call r , at most $d+1$ vectors $x^{(r,t)}$ are generated. Hence, matrix A has at most $d \cdot (d+1)$ columns and at most $d^3 + d^2 + d$ rows. This yields the claimed bound

$$\begin{aligned}
|\mathcal{C}| &\leq d^d \cdot (d+1)^d \cdot d^2 \cdot (d+1) \cdot n^{d^2(d+1)} \cdot (\mathcal{K} + 1)^{d(d+1) \cdot (d^3+d^2+d)} \\
&\leq 2^{d+1} \cdot d^{2d+3} \cdot n^{d^3+d^2} \cdot (\mathcal{K} + 1)^{(d^2+d)(d^3+d^2+d)} \\
&= (\mathcal{K} + 1)^{(d^2+d)(d^3+d^2+d)} \cdot O(n^{d^3+d^2}). \quad \square
\end{aligned}$$

In the next step we analyze how much information of V is required in order to perform the call $\text{Witness}_0(V, [d], 1, I^*, A, \mathcal{S}, B, u)$ for a fixed certificate (I^*, A) . We will see that V does not need to be revealed completely and that some randomness remains even after the necessary information to perform the call has been revealed. This is the key observation for analyzing the probability $\mathbf{Pr}_V[E_{I^*, A, B}]$. For this, let V be an arbitrary realization, i.e., we do not condition on the OK_0 -event anymore. Let $I_k^* = I^* \cap P_k$ for $k \in [d]$. We apply the principle of deferred decisions and

assume that for every $k \in [d]$ the coefficients of V^k belonging to indices $i \notin I_k^*$ are fixed arbitrarily. We denote this part of V^k by $V_{I_k^*}^k$ and concentrate on the remaining part of V^k which we denote by $V_{I_k^*}^k$.

As in the model with non-zero-preserving perturbations, the call $\text{Witness}_0(V, [d], 1, I^*, A, \mathcal{S}, B, u)$ can be executed without the full knowledge of $V_{I_1^*}^1, \dots, V_{I_d^*}^d$. We write the linear combinations of $V_{I_k^*}^k$ of the calls $r = 1, \dots, r_k - 1$ and of call r_k that suffice to be known into the following matrices P_k and Q_k , respectively:

$$P_k = \left[p_k^{(1, d_1)}, \dots, p_k^{(1, t_1)}, \dots, p_k^{(r_k-1, d_{r_k-1})}, \dots, p_k^{(r_k-1, t_{r_k-1})} \right] \Big|_{I_k^*} \quad \text{and}$$

$$Q_k = \left[p_k^{(r_k, d_{r_k})}, \dots, p_k^{(r_k, j_k)}, p_k^{(r_k, j_k-2)} - p_k^{(r_k, j_k-1)}, \dots, p_k^{(r_k, t_{r_k})} - p_k^{(r_k, j_k-1)} \right] \Big|_{I_k^*}$$

for $p_k^{(r, t)} = a^{(r, t)}|_{I_k^*} - u|_{I_k^*}$, where $a^{(r, t)}$ are the columns of matrix A . The index $j_k \in [d_{r_k}]$ denotes the index for which $k_{j_k} = k$ in the r_k^{th} call of the Witness_0 function, i.e., V^k is the j_k^{th} objective function in K in the call $r = r_k$ and it is not considered anymore in iterations $t < j_k$.

Note that the matrices $P_k = P_k(I^*, A, u)$ and $Q_k = Q_k(I^*, A, u)$ depend, among others, on the choice of u . Furthermore, the indices j_k are determined by the certificate (I^*, A) . To be precise, the indices r_1, \dots, r_d , which are implicitly given by the certificate (I^*, A) , contain the information which objectives are still under consideration in a certain call r of the Witness_0 function: These are all objectives V^k for which $r_k \geq r$.

Observe that matrix Q_k has

$$(d_{r_k} - j_k + 1) + ((j_k - 2) - t_{r_k} + 1) = d_{r_k} - t_{r_k}$$

columns and matrix P_k has

$$\sum_{r=1}^{r_k-1} (d_r - t_r + 1) = \sum_{r=1}^{r_k} (d_r - t_r + 1) - (d_{r_k} - t_{r_k} + 1) = |I_k^*| - (d_{r_k} - t_{r_k} + 1)$$

columns. The last equation is due to the fact that in each call $r < r_k$ in each iteration one index i_k is chosen. In call $r = r_k$ one index i_k is chosen in each iteration $t > t_{r_k}$. The equation follows since I_k^* contains one index more than the number of indices i_k that are chosen during the execution of the Witness_0 function (see Definition 31). Moreover, observe that all entries of P_k and Q_k are from $\{-\mathcal{K}, \dots, \mathcal{K}\}$.

Lemma 37. *Let $u \in \{0, \dots, \mathcal{K}\}^n$ be an arbitrary shift vector, let $(I^*, A) \in \mathcal{C}$ be an arbitrary certificate, and let U and W be two realizations for which $U_{I_k^*}^k = W_{I_k^*}^k$ and $U_{I_k^*}^k \cdot q = W_{I_k^*}^k \cdot q$ for all indices $k \in [d]$ and all columns q of one of the matrices $P_k(I^*, A, u)$ and $Q_k(I^*, A, u)$. Then for every ε -box $B \in \mathbb{B}_\varepsilon$ the calls $\text{Witness}_0(U, [d], 1, I^*, A, \mathcal{S}, B, u)$ and $\text{Witness}_0(W, [d], 1, I^*, A, \mathcal{S}, B, u)$ return the same result.*

Proof. We fix an index $k \in [d]$ and analyze which information of $V_{I_k^*}^k$ is required for the execution of the call $\text{Witness}_0(V, [d], 1, I^*, A, \mathcal{S}, B, u)$. By the construction of index r_k we know that only in the calls $r = 1, \dots, r_k$ information about the function V^k must be available as in all subsequent calls this function is not considered anymore.

Since in call r we consider only vectors that coincide with one of the vectors $x^{(r, t_r)}, \dots, x^{(r, d_r)}$ (see Line 5) in the indices $i \in I^*$, it suffices to know all linear combinations

$$V_{I_k^*}^k \cdot \left(a^{(r, t)}|_{I_k^*} - u|_{I_k^*} \right) = V_{I_k^*}^k \cdot p_k^{(r, t)}.$$

For all call numbers $r = 1, \dots, r_k - 1$ and all iterations $t = t_r, \dots, d_r$ in these calls the vector $p_k^{(r, t)}$ is a column of matrix P_k .

It remains to analyze which information of $V_{I_k^*}^k$ is required in the r_k^{th} call of the `Witness0` function. First, we observe that $t_{r_k} \leq j_k - 1$. This is due to the fact that $x^{(r_k, t_{r_k})}|_{P_k} = x|_{P_k}$, i.e., $V^k x^{(r_k, t_{r_k})} = V^k x$, since this is the iteration when k is removed from tuple K . On the other hand, $x^{(r_k, t_{r_k})} \in \mathcal{C}_{t_{r_k}}^{(r_k)}$, which means that $V^{k_s} x^{(r_k, t_{r_k})} < V^{k_s} x$ for all indices $s = 1, \dots, t_{r_k}$, where $k_1, \dots, k_{d_{r_k}}$ denote the indices tuple K consists of in call $r = r_k$. Hence, $t_{r_k} < j_k$ as $k = k_{j_k}$.

There are only three lines where information about V is required: Line 7, Line 9, and Line 15. For Line 7 the values

$$V_{I_k^*}^k \cdot \left(a^{(r_k, t)}|_{I_k^*} - u|_{I_k^*} \right) = V_{I_k^*}^k \cdot p_k^{(r_k, t)}$$

from iteration $t = d_{r_k}$ down to iteration j_k are needed. The vectors $p^{(r_k, j_k)}, \dots, p^{(r_k, d_{r_k})}$ are columns of matrix Q_k . In Line 9 no additional information about $V_{I_k^*}^k$ is required since all considered vectors agree on indices $i \in I_k^*$ with each other. For Line 15 only in iteration $t = j_k - 1$ the values

$$V_{I_k^*}^k \cdot \left(a^{(r_k, s)}|_{I_k^*} - a^{(r_k, j_k - 1)}|_{I_k^*} \right)$$

for $s = t_{r_k}, \dots, j_k - 2$ are required. Observe that the vectors

$$\begin{aligned} p_k^{(r_k, s)} - p_k^{(r_k, j_k - 1)} &= \left(a^{(r_k, s)}|_{I_k^*} - u|_{I_k^*} \right) - \left(a^{(r_k, j_k - 1)}|_{I_k^*} - u|_{I_k^*} \right) \\ &= a^{(r_k, s)}|_{I_k^*} - a^{(r_k, j_k - 1)}|_{I_k^*} \end{aligned}$$

for $s = t_{r_k}, \dots, j_k - 2$ are columns of matrix Q_k .

As U and W agree on all necessary information, both calls return the same result. \square

In the remainder of this section we assume that $V_{I_k^*}^k$ and the ε -box B are fixed. In accordance with Lemma 37, the output of the call `Witness0`($V, [d], 1, I^*, A, \mathcal{S}, B, u$) is determined if the linear combinations of $V_{I_k^*}^k$ given by the columns of the matrices P_k and Q_k are fixed arbitrarily, i.e., it does not depend on the remaining randomness in the coefficients. We are interested in the event $E_{I^*, A, B}$, i.e., in the event that the output is a set $\{x'\}$ such that $V^{1 \dots d} \cdot (x' - u^*(I^*, A)) \in B$. Since (I^*, A) is a V -certificate of x for some V and x , the output is always of the form $\{x'\}$ due to Lemma 33. Hence, event $E_{I^*, A, B}$ occurs if and only if for all indices k the relation $V_{I_k^*}^k \cdot (x' - u^*(I^*, A))|_{I_k^*} \in C_k$ holds for some interval C_k of length ε that depends on the linear combinations of $V_{I_\ell^*}$ given by P_ℓ and Q_ℓ for all indices $\ell \in [d]$.

Lemma 38. *For every fixed index $k \in [d]$ the columns of matrix $P_k(I^*, A, u^*(I^*, A))$, of matrix $Q_k(I^*, A, u^*(I^*, A))$, and the vector $p_k^{(r_k, t_{r_k})}$ are linearly independent.*

Proof. Consider the square matrix \hat{Q}_k consisting of the vectors $p_k^{(r, t)}$, for $r \in \{1, \dots, r_k\}$ and $t \in \{t_r, \dots, d_r\}$. Matrix \hat{Q}_k can be obtained from the matrix M of Lemma 32 by subtracting the vector $u^*|_{I_k^*}$ from all of its columns. Due to Lemma 32 and due to the construction of $u^* = u^*(I^*, A)$ (see Equation 3) matrix \hat{Q}_k is a lower triangular matrix and the elements of the principal diagonal are from the set $\{-\mathcal{K}, \dots, \mathcal{K}\} \setminus \{0\}$. This is because $x_i - u_i^* = x_i - |x_i - 1| \in \{-1, 1\}$ for $i = i_k^*$ and $\bar{x}_i - u_i^* = \bar{x}_i - x_i \neq 0$ for all $i \in I_k^* \setminus \{i_k^*\}$, where \bar{z} for $z \in \{0, \dots, \mathcal{K}\}$ represents an arbitrary value from $\{0, \dots, \mathcal{K}\}$ not equal to z . Hence, the vectors $p_k^{(r, t)}$ are linearly independent.

The columns of Q_k and $p_k^{(r_k, t_{r_k})}$ are linear combinations of the vectors $p_k^{(r_k, t_{r_k})}, \dots, p_k^{(r_k, d_{r_k})}$, whereas the columns of matrix P_k are the remaining columns of matrix \hat{Q}_k . As the vectors $p_k^{(r, t)}$ are linearly independent, it suffices to show that the columns of matrix Q_k and vector $p_k^{(r_k, t_{r_k})}$ are linearly independent. For this, we consider an arbitrary linear combination of the columns of matrix Q_k and vector $p_k^{(r_k, t_{r_k})}$ and show that it is 0 if and only if all coefficients are 0. For sake

of simplicity, we drop the index k in the remainder of this proof and write r , j , and $p^{(r,t)}$ instead of r_k , j_k , and $p_k^{(r_k,t)}$, respectively.

$$\sum_{t=j}^{d_r} \lambda_t \cdot p^{(r,t)} + \sum_{t=t_r}^{j-2} \lambda_t \cdot \left(p^{(r,t)} - p^{(r,j-1)} \right) + \mu \cdot p^{(r,t_r)} = 0.$$

If $t_r = j - 1$, then this equation is equivalent to

$$\sum_{t=t_r+1}^{d_r} \lambda_t \cdot p^{(r,t)} + \mu \cdot p^{(r,t_r)} = 0.$$

Therefore, all coefficients are 0 due to the linear independence of the vectors $p^{(r,t)}$. If $t_r < j - 1$, which is the only case remaining due to previous observations, then the equation is equivalent to

$$\sum_{t=j}^{d_r} \lambda_t \cdot p^{(r,t)} + \sum_{t=t_r+1}^{j-2} \lambda_t \cdot p^{(r,t)} - \left(\sum_{t=t_r}^{j-2} \lambda_t \right) \cdot p^{(r,j-1)} + (\lambda_{t_r} + \mu) \cdot p^{(r,t_r)} = 0.$$

The linear independence of the vectors $p^{(r,t)}$ implies $\lambda_t = 0$ for $t \in \{t_r + 1, \dots, j - 2\} \cup \{j, \dots, d_r\}$, $\sum_{t=t_r}^{j-2} \lambda_t = 0$, and $\lambda_{t_r} + \mu = 0$. Consequently, also $\lambda_{t_r} = 0$ and, thus, $\mu = 0$. This means that all coefficients are 0. In both cases, the linear independence of the columns of Q_k and the vector $p^{(r,t_r)}$ follows. \square

Corollary 39. *Let $\gamma = d^3 + d^2 + d$. For an arbitrary certificate $(I^*, A) \in \mathcal{C}$ the probability of the event $E_{I^*, A, B}$ is bounded by*

$$\Pr_V [E_{I^*, A, B}] \leq (2\gamma\mathcal{K})^{\gamma-d} \phi^\gamma \varepsilon^d$$

and by

$$\Pr_V [E_{I^*, A, B}] \leq 2^d (\gamma\mathcal{K})^{\gamma-d} \phi^d \varepsilon^d$$

if all densities are quasiconcave.

Proof. For a fixed index $k \in [d]$ we write the columns of matrix P_k , of matrix Q_k , and vector $p_k^{(r_k, t_{r_k})}$ into one matrix $Q'_k \in \{-\mathcal{K}, \dots, \mathcal{K}\}^{|I_k^*| \times |I_k^*|}$ (the number of columns is

$$\left(|I_k^*| - (d_{r_k} - t_{r_k} + 1) \right) + (d_{r_k} - t_{r_k}) + 1 = |I_k^*|$$

due to previous observations) and consider the matrix

$$Q' = \begin{bmatrix} Q'_1 & \mathbb{O} & \dots & \mathbb{O} \\ \mathbb{O} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbb{O} \\ \mathbb{O} & \dots & \mathbb{O} & Q'_d \end{bmatrix} \in \{-\mathcal{K}, \dots, \mathcal{K}\}^{|I^*| \times |I^*|}.$$

This matrix has full rank due to Lemma 38. Now we permute the columns of Q' to obtain a matrix Q whose last d columns belong to the last column of one of the matrices Q_k . This means that the last d columns are

$$\left(p_1^{(r_1, t_{r_1})}, \mathbb{O}^{|I_2^*|}, \dots, \mathbb{O}^{|I_d^*|} \right), \dots, \left(\mathbb{O}^{|I_1^*|}, \dots, \mathbb{O}^{|I_{d-1}^*|}, p_d^{(r_d, t_{r_d})} \right).$$

For all $k \in [d]$ and every index $i_k \in I_k^*$ let $X_i = V_i^k$ be the i^{th} coefficient of V^k . Event $E_{I^*, A, B}$ holds if and only if the d linear combinations of the variables X_i given by the last d columns of Q fall into a d -dimensional hypercube C depending on the linear combinations of the variables X_i given by the remaining columns. The claim follows by applying Theorem 40 for matrix $A = Q'^T$ and due to the fact that $|I^*| \leq \gamma$ (see proof of Lemma 36). \square

Proof of Theorem 1. We begin the proof by showing that the OK_0 -event is likely to happen. For all indices $t \in [d]$ and all solutions $x, y \in \mathcal{S}$ for which $x|_{P_t} \neq y|_{P_t}$ the probability that $|V^t x - V^t y| \leq \varepsilon$ is bounded by $2\phi\varepsilon$. To see this, choose one index $i \in P_t$ for which $x_i \neq y_i$ and apply the principle of deferred decisions by fixing all coefficients V_j^t for $j \neq i$ arbitrarily. Then the value V_i^t must fall into an interval of length $2\varepsilon/|x_i - y_i| \leq 2\varepsilon$. The probability for this is bounded by $2\phi\varepsilon$. A union bound over all indices $t \in [d]$ and over all pairs $(x, y) \in \mathcal{S} \times \mathcal{S}$ for which $x|_{P_t} \neq y|_{P_t}$ yields

$$\Pr_V \left[\overline{\text{OK}_0(V)} \right] \leq 2(\mathcal{K} + 1)^{2n} d\phi\varepsilon.$$

Let $\gamma = d^3 + d^2 + d$. We set

$$s = \begin{cases} (2\gamma\mathcal{K})^{\gamma-d}\phi^\gamma = \mathcal{K}^{\gamma-d} \cdot O(\phi^\gamma) & \text{for general density functions,} \\ 2^d(\gamma\mathcal{K})^{\gamma-d}\phi^d = \mathcal{K}^{\gamma-d} \cdot O(\phi^d) & \text{for quasiconcave density functions.} \end{cases}$$

Then we obtain

$$\begin{aligned} \mathbf{E}_V[\text{PO}(V)] &\leq \sum_{(I^*, A) \in \mathcal{C}} \sum_{B \in \mathbb{B}_\varepsilon} \Pr_V[E_{I^*, A, B}] + (\mathcal{K} + 1)^n \cdot \Pr_V \left[\overline{\text{OK}_0(V)} \right] \\ &\leq \sum_{(I^*, A) \in \mathcal{C}} \sum_{B \in \mathbb{B}_\varepsilon} s \cdot \varepsilon^d + (\mathcal{K} + 1)^n \cdot 2(\mathcal{K} + 1)^{2n} d\phi\varepsilon \\ &= |\mathcal{C}| \cdot |\mathbb{B}_\varepsilon| \cdot s \cdot \varepsilon^d + (\mathcal{K} + 1)^n \cdot 2(\mathcal{K} + 1)^{2n} d\phi\varepsilon \\ &= (\mathcal{K} + 1)^{(d^2+d)(d^3+d^2+d)} \cdot O(n^{d^3+d^2}) \cdot \left(\frac{2n\mathcal{K}}{\varepsilon} \right)^d \cdot s \cdot \varepsilon^d + 2(\mathcal{K} + 1)^{3n} d\phi\varepsilon \\ &= \mathcal{K}^{(d^2+d)(d^3+d^2+d)+d} \cdot O(n^{d^3+d^2+d}) \cdot s + 2(\mathcal{K} + 1)^{3n} d\phi\varepsilon. \end{aligned}$$

The first inequality is due to Corollary 35. The second inequality is due to Corollary 39. The third inequality stems from Lemma 36. Since this bound holds for any $\varepsilon > 0$ for which $1/\varepsilon$ is integral, it also holds for the limit $\varepsilon \rightarrow 0$. Hence, we obtain

$$\mathbf{E}_V[\text{PO}(V)] = \mathcal{K}^{(d^2+d)(d^3+d^2+d)+d} \cdot O(n^{d^3+d^2+d}) \cdot s.$$

Substituting s and γ by their definitions yields

$$\begin{aligned} \mathbf{E}_V[\text{PO}(V)] &= \mathcal{K}^{(d^2+d)(d^3+d^2+d)+d} \cdot O(n^{d^3+d^2+d}) \cdot \mathcal{K}^{d^3+d^2+d-d} \cdot O(\phi^{d^3+d^2+d}) \\ &= \mathcal{K}^{(d^2+d+1)(d^3+d^2+d)} \cdot O((n\phi)^{d^3+d^2+d}) \\ &\leq \mathcal{K}^{(d+1)^5} \cdot O((n\phi)^{d^3+d^2+d}) \end{aligned}$$

for general densities and

$$\begin{aligned} \mathbf{E}_V[\text{PO}(V)] &= \mathcal{K}^{(d^2+d)(d^3+d^2+d)+d} \cdot O(n^{d^3+d^2+d}) \cdot \mathcal{K}^{\gamma-d} \cdot O(\phi^d) \\ &= \mathcal{K}^{(d^2+d+1)(d^3+d^2+d)} \cdot O(n^{d^3+d^2+d} \phi^d) \\ &\leq \mathcal{K}^{(d+1)^5} \cdot O(n^{d^3+d^2+d} \phi^d) \end{aligned}$$

for quasiconcave densities. □

7 Some Probability Theory

In this chapter we lay the probabilistic foundation of this article. We consider linearly independent linear combinations of independent random variables and show that they behave to some extent like independent random variables.

Let X_1, \dots, X_n be independent random variables with densities $f_i: [-1, 1] \rightarrow [0, \phi]$ for all $i \in [n]$ and let $A \in \{-\mathcal{K}, \dots, \mathcal{K}\}^{m \times n}$ be an integer matrix. Furthermore, let $(Y_1, \dots, Y_{m-k}, Z_1, \dots, Z_k)^\top = A \cdot (X_1, \dots, X_n)^\top$ be an m -dimensional random vector whose entries are linear combinations of the random variables X_1, \dots, X_n , and let C be an arbitrary function that maps every tuple $(y_1, \dots, y_{m-k}) \in \mathbb{R}^{m-k}$ to a k -dimensional hypercube $C(y_1, \dots, y_{m-k}) \subseteq \mathbb{R}^k$ with side length ε . We want to bound the probability that the random vector $(Z_1, \dots, Z_k)^\top$ falls into the random hypercube $C(Y_1, \dots, Y_{m-k})$ from above.

Before we state the main theorem of this section, let us discuss two special cases. One simple case is when the matrix A is of the form $A = [\mathbb{I}_m, \mathbb{O}_{m \times n-m}]$. In this case, the random variables $Y_i = X_i$, $i = 1, \dots, m-k$, and $Z_j = X_{m-k+j}$, $j = 1, \dots, k$, are independent. In order to bound the probability $\Pr[(Z_1, \dots, Z_k) \in C(Y_1, \dots, Y_{m-k})]$, we can apply the principle of deferred decisions and assume that the outcome of the variables Y_1, \dots, Y_{m-k} has been revealed by an adversary, say $Y_i = y_i$ for $i = 1, \dots, m-k$. Hence, the hypercube $C(Y_1, \dots, Y_{m-k}) = C(y_1, \dots, y_{m-k})$ is fixed and not random anymore. However, we still have not revealed the outcome of the random variables Z_1, \dots, Z_k . As the random variables $Y_1, \dots, Y_{m-k}, Z_1, \dots, Z_k$ are independent, we obtain

$$\begin{aligned} \Pr[(Z_1, \dots, Z_k) \in C(Y_1, \dots, Y_{m-k}) \mid (Y_1, \dots, Y_{m-k}) = (y_1, \dots, y_{m-k})] \\ &= \Pr[(Z_1, \dots, Z_k) \in C(y_1, \dots, y_{m-k})] \\ &= \Pr[(X_{m-k+1}, \dots, X_m) \in C(y_1, \dots, y_{m-k})] \\ &\leq (\phi\varepsilon)^k. \end{aligned}$$

Observe that we used the simple structure of A twice: First, we obtained independence which is why the first equation holds. Second, the event $(Z_1, \dots, Z_k) \in C(y_1, \dots, y_{m-k})$ can be directly translated into the event $(X_{m-k+1}, \dots, X_m) \in C(y_1, \dots, y_{m-k})$ that only depends on the k random variables X_{m-k+1}, \dots, X_m and the hypercube $C(y_1, \dots, y_{m-k})$. In general, all random variables X_1, \dots, X_n and a more complex set $\tilde{C}(y_1, \dots, y_{m-k})$ that depends on the last k rows of A have to be considered.

Now let us consider a second special case in which each of the last k rows $a_{m-k+1}^\top, \dots, a_m^\top$ of A is a linear combination of the first $m-k$ rows $a_1^\top, \dots, a_{m-k}^\top$. In particular, for every index $i = 1, \dots, k$ we can write a_{m-k+i} as

$$a_{m-k+i} = \sum_{j=1}^{m-k} \lambda_j^{(i)} \cdot a_j$$

for appropriate coefficients $\lambda_j^{(i)}$. As the function C we consider the function that maps a tuple (y_1, \dots, y_{m-k}) to the hypercube $[b_1, b_1 + \varepsilon] \times \dots \times [b_k, b_k + \varepsilon]$, where b_i is defined as

$$b_i = \sum_{j=1}^{m-k} \lambda_j^{(i)} \cdot y_j.$$

With this choice and the notation $X = (X_1, \dots, X_n)^\top$ we obtain

$$Z_i = a_{m-k+i}^\top X = \sum_{j=1}^{m-k} \lambda_j^{(i)} \cdot a_j^\top X = \sum_{j=1}^{m-k} \lambda_j^{(i)} \cdot Y_j$$

for all $i = 1, \dots, k$. Hence, (Z_1, \dots, Z_k) falls into the hypercube $C(Y_1, \dots, Y_{m-k})$ for every realization of the random variables X_1, \dots, X_n . Consequently, $\Pr[(Z_1, \dots, Z_k) \in C(Y_1, \dots, Y_{m-k})] = 1$. The reason why we can define a function C with such a property is that the outcome of the random variables Z_1, \dots, Z_k is determined when the outcome of the random variables Y_1, \dots, Y_{m-k} has been revealed since the last k rows of A can be expressed as linear combinations of the first $m-k$ rows of A . Hence, in this special case we cannot obtain any non-trivial bound.

The following theorem claims a non-trivial bound for the probability that the random vector (Z_1, \dots, Z_k) falls into the hypercube $C(Y_1, \dots, Y_{m-k})$ for the case when the rows of matrix A are linearly independent. The first inequality of Theorem 40 has been shown for binary matrices by Röglin and Teng [18] (see Lemma 3.3). Their proof can be easily generalized to arbitrary integer matrices. For the sake of completeness we state it here.

Theorem 40. *Let $m \leq n$ be integers and let X_1, \dots, X_n be independent random variables, each with a probability density function $f_i: [-1, 1] \rightarrow [0, \phi]$, let $A \in \{-\mathcal{K}, \dots, \mathcal{K}\}^{m \times n}$ be a matrix of rank m , let $k \in [m-1]$ be an integer, let*

$$(Y_1, \dots, Y_{m-k}, Z_1, \dots, Z_k)^T = A \cdot (X_1, \dots, X_n)^T$$

be the linear combinations of X_1, \dots, X_n given by A , and let C be a function mapping a tuple $(y_1, \dots, y_{m-k}) \in \mathbb{R}^{m-k}$ to a hypercube $C(y_1, \dots, y_{m-k}) \subseteq \mathbb{R}^k$ with side length ε . Then

$$\Pr[(Z_1, \dots, Z_k) \in C(Y_1, \dots, Y_{m-k})] \leq (2m\mathcal{K})^{m-k} \phi^m \varepsilon^k.$$

If all densities f_i are quasiconcave, then even the stronger bound

$$\Pr[(Z_1, \dots, Z_k) \in C(Y_1, \dots, Y_{m-k})] \leq 2^k (m\mathcal{K})^{m-k} \phi^k \varepsilon^k$$

holds.

Theorem 40 states that, for quasiconcave densities, linearly independent linear combinations of independent random variables almost behave like independent random variables when considering the event $(Z_1, \dots, Z_k) \in C(Y_1, \dots, Y_{m-k})$ with respect to ϕ and ε : The bound in Theorem 40 deviates from the bound derived for the special case $Y_i = X_i$ for $i = 1, \dots, m-k$ and $Z_j = X_{m-k+j}$ for $j = 1, \dots, k$ (see beginning of this section) only by a factor of $2^k (m\mathcal{K})^{m-k}$.

Proof. First of all we show that we can assume w.l.o.g. that $n = m$. Otherwise, we can choose m indices $i_1 < \dots < i_m \in [n]$ for which the matrix $A' = [a_{i_1}, \dots, a_{i_m}]$ is a full-rank square submatrix of A . For the sake of simplicity let us assume that $i_k = k$ for $k = 1, \dots, m$. We apply the principle of deferred decisions and assume that X_{m+1}, \dots, X_n are fixed arbitrarily to some values x_{m+1}, \dots, x_n .

Let $A'' = [a_{m+1}, \dots, a_n]$, $A'_1 = A''|_{1, \dots, m-k}$, $A'_2 = A''|_{m-k+1, \dots, m}$, and $x = (x_{m+1}, \dots, x_n)$. Let us further introduce the random vector

$$(Y'_1, \dots, Y'_{m-k}, Z'_1, \dots, Z'_k) = A' \cdot (X_1, \dots, X_m)$$

and the function

$$C'(Y'_1, \dots, Y'_{m-k}) = C((Y'_1, \dots, Y'_{m-k}) + A'_1 \cdot x) - A'_2 \cdot x.$$

Observing that

$$\begin{aligned} (Y'_1, \dots, Y'_{m-k}) + A'_1 \cdot x &= (Y_1, \dots, Y_{m-k}) \quad \text{and} \\ (Z'_1, \dots, Z'_k) + A'_2 \cdot x &= (Z_1, \dots, Z_k), \end{aligned}$$

we obtain

$$\begin{aligned} (Z_1, \dots, Z_k) \in C(Y_1, \dots, Y_{m-k}) &\iff (Z'_1, \dots, Z'_k) \in C(Y_1, \dots, Y_{m-k}) - A'_2 \cdot x \\ &\iff (Z'_1, \dots, Z'_k) \in C'(Y'_1, \dots, Y'_{m-k}). \end{aligned}$$

The probability of the last event can be bounded by applying Theorem 40 for the $m \times m$ -matrix A' .

In the remainder of this proof we assume that $n = m$. As matrix A is a full-rank square matrix, its inverse A^{-1} exists and we can write

$$\begin{aligned} \Pr[(Z_1, \dots, Z_k) \in C(Y_1, \dots, Y_{n-k})] &= \int_{y \in \mathbb{R}^{n-k}} \int_{z \in C(y)} f_{Y,Z}(y, z) \, dz \, dy \\ &= \int_{y \in \mathbb{R}^{n-k}} \int_{z \in C(y)} |\det(A^{-1})| \cdot f_X(A^{-1} \cdot (y, z)) \, dz \, dy \\ &\leq \int_{y \in \mathbb{R}^{n-k}} \int_{z \in C(y)} f_X(A^{-1} \cdot (y, z)) \, dz \, dy \\ &\leq \varepsilon^k \cdot \int_{y \in \mathbb{R}^{n-k}} \max_{z \in \mathbb{R}^k} f_X(A^{-1} \cdot (y, z)) \, dy, \end{aligned}$$

where $f_{Y,Z}$ denotes the common density of the variables $Y_1, \dots, Y_{n-k}, Z_1, \dots, Z_k$ and $f_X = \prod_{i=1}^n f_i$ denotes the common density of the variables X_1, \dots, X_n . The second equality is due to a change of variables, the first inequality stems from the fact that $|\det(A^{-1})| = |1/\det A| \leq 1$ since A is an integer matrix.

In general, we can bound the integral in the formula above by

$$\begin{aligned} \int_{y \in \mathbb{R}^{n-k}} \max_{z \in \mathbb{R}^k} f_X(A^{-1} \cdot (y, z)) \, dy &\leq \int_{y \in [-n\mathcal{K}, n\mathcal{K}]^{n-k}} \max_{z \in \mathbb{R}^k} f_X(A^{-1} \cdot (y, z)) \, dy \\ &\leq \int_{y \in [-n\mathcal{K}, n\mathcal{K}]^{n-k}} \phi^n \, dy \\ &= (2n\mathcal{K})^{n-k} \phi^n \\ &= (2m\mathcal{K})^{m-k} \phi^m, \end{aligned}$$

where the first inequality is due to the fact that all variables Y_i can only take values in the interval $[-n\mathcal{K}, n\mathcal{K}]$ as all entries of matrix A are from $\{-\mathcal{K}, \dots, \mathcal{K}\}$ and as all variables X_j can only take values in the interval $[-1, 1]$.

To prove the statement about quasiconcave functions we first consider arbitrary rectangular functions, i.e., functions that are constant on a given interval, and 0 otherwise. This will be the main part of our analysis. Afterwards, we analyze sums of rectangular functions and, finally, we show that quasiconcave functions can be approximated by such sums.

Lemma 41. *For $i \in [n]$ let $\phi_i \geq 0$, let $I_i \subseteq \mathbb{R}$ be an interval of length ℓ_i , and let $f_i: \mathbb{R} \rightarrow \mathbb{R}$ be the function*

$$f_i(x) = \begin{cases} \phi_i & \text{if } x \in I_i, \\ 0 & \text{otherwise.} \end{cases}$$

Moreover, let $f: \mathbb{R}^n \rightarrow \mathbb{R}$ be the function $f(x_1, \dots, x_n) = \prod_{i=1}^n f_i(x_i)$ and let $A \in \{-\mathcal{K}, \dots, \mathcal{K}\}^{n \times n}$ be an invertible matrix. Then

$$\int_{y \in \mathbb{R}^{n-k}} \max_{z \in \mathbb{R}^k} f(A^{-1} \cdot (y, z)) \, dy \leq 2^k \cdot (n-k)! \cdot \mathcal{K}^{n-k} \cdot \chi \cdot \sum_I \prod_{i \notin I} \ell_i$$

where $\chi = \prod_{i=1}^n \phi_i$ and where the sum runs over all tuples $I = (i_1, \dots, i_k)$ for which $1 \leq i_1 < \dots < i_k \leq n$.

Proof. Function f takes the value χ on the n -dimensional box $Q = \prod_{i=1}^n I_i$ and is 0 otherwise. Hence,

$$\int_{y \in \mathbb{R}^{n-k}} \max_{z \in \mathbb{R}^k} f(A^{-1} \cdot (y, z)) dy = \chi \cdot \text{vol}(Q')$$

for

$$\begin{aligned} Q' &= \{y \in \mathbb{R}^{n-k} : \exists z \in \mathbb{R}^k \text{ such that } A^{-1} \cdot (y, z) \in Q\} \\ &= \{y \in \mathbb{R}^{n-k} : \exists z \in \mathbb{R}^k \exists x \in Q \text{ such that } (y, z) = A \cdot x\} \\ &= (P \cdot A)(Q), \end{aligned}$$

where $P := [\mathbb{I}_{n-k}, \mathbb{O}_{(n-k) \times k}]$ is the projection matrix that removes the last k entries from a vector of length n . In the remainder of this proof we bound the volume of $M(Q)$ where $M := P \cdot A \in \{-\mathcal{K}, \dots, \mathcal{K}\}^{(n-k) \times n}$. Let $a_i := c_i^0$ and $b_i := c_i^1$ be the left and the right bound of interval I_i , respectively. For an index tuple $I = (i_1, \dots, i_k)$, $1 \leq i_1 < \dots < i_k \leq n$, and a bit tuple $J = (j_1, \dots, j_k) \in \{0, 1\}^k$, let

$$F_I^J = \prod_{i=1}^n \begin{cases} \{c_{i_t}^{j_t}\} & \text{if } i = i_t \in I, \\ I_i & \text{if } i \notin I, \end{cases}$$

be one of the $2^k \cdot \binom{n}{k}$ $(n-k)$ -dimensional faces of Q . We show that $M(Q) \subseteq \bigcup_I \bigcup_J M(F_I^J)$. Let $y \in M(Q)$, i.e., there is a vector $x \in Q$ such that $y = M \cdot x$. Now, consider the polytope

$$R = \{(x', s') \in \mathbb{R}^n \times \mathbb{R}^n : M \cdot x' = y', \ x' + s' = b', \ \text{and } x', s' \geq 0\},$$

where $y' = y - M \cdot a$ and $b' = b - a$ for $a = (a_1, \dots, a_n)$ and $b = (b_1, \dots, b_n)$. This polytope is bounded and non-empty because $(x - a, b - x) \in R$. Consequently, there exists a basic feasible solution (x^*, s^*) . As there are $2n$ variables and $2n - k$ constraints, this solution has at least k zero-entries, i.e., there are indices $1 \leq i_1 < \dots < i_k \leq n$ such that either $x_{i_t}^* = 0$ (in that case set $j_t = 0$) or $x_{i_t}^* = b_{i_t}'$ (in that case set $j_t = 1$) for all $t \in [k]$. Now, consider the vector $\hat{x} = x^* + a \in [0, b'] + a = Q$. We obtain $M \cdot \hat{x} = y$ and $\hat{x}_{i_t} = c_{i_t}^{j_t}$ for all $t \in [k]$. Hence, $x \in F_I^J$ for $I = (i_1, \dots, i_k)$ and $J = (j_1, \dots, j_k)$, and thus $y \in M(F_I^J)$.

Due to this observation we can bound the volume of $M(Q)$ by $\sum_I \sum_J \text{vol}(M(F_I^J))$. It remains to show how to bound the volume $\text{vol}(M(F_I^J))$. For the sake of simplicity we only consider $I = (n - k + 1, \dots, n)$ in the following analysis. Let $\phi^J: \mathbb{R}^{n-k} \rightarrow F_I^J$ be the function $\phi^J(x) = T \cdot x + v^J$, where $T = [\mathbb{I}_{n-k}, \mathbb{O}_{(n-k) \times k}]^T$ and $v^J = (0, \dots, 0, c_{n-k+1}^{j_1}, \dots, c_n^{j_k})$. Using function ϕ^J is the canonical way to describe the affine subspace defined by face F_I^J : it adds the fixed coordinates of F_I^J to a given vector of length $n-k$. Hence, function ϕ^J , restricted to the domain $F' = \prod_{i=1}^{n-k} I_i$, is bijective. With $\psi = M \circ \phi^J$ we obtain

$$\begin{aligned} \text{vol}(M(F_I^J)) &= \int_{\psi(F')} 1 dx = \int_{F'} |\det D\psi(x)| dx = \int_{F'} |\det(M \cdot T)| dx \\ &= |\det(M \cdot T)| \cdot \text{vol}(F') = |\det(M \cdot T)| \cdot \prod_{i=1}^{n-k} \ell_i. \end{aligned}$$

In general, the second equality only holds if ψ is injective. If ψ is not injective, then $M(F_I^J) = \psi(F')$ is not full-dimensional, i.e., $\text{vol}(M(F_I^J)) = 0$, and $\det(M \cdot T) = 0$ since ψ is affine linear. Hence, the second equality also holds in the case when ψ is not injective.

Matrix $M \cdot T = P \cdot A \cdot T$ is an $(n-k) \times (n-k)$ -submatrix of A . Thus, $|\det(M \cdot T)| \leq (n-k)! \cdot \mathcal{K}^{n-k}$, and we obtain the bound

$$\int_{y \in \mathbb{R}^{n-k}} \max_{z \in \mathbb{R}^k} f(A^{-1} \cdot (y, z)) dy = \chi \cdot \text{vol}(M(Q)) \leq \chi \cdot \sum_I \sum_J \text{vol}(M(F_I^J))$$

$$\begin{aligned}
&\leq \chi \cdot \sum_I \sum_J (n-k)! \cdot \mathcal{K}^{n-k} \cdot \prod_{i \notin I} \ell_i \\
&= 2^k \cdot (n-k)! \cdot \mathcal{K}^{n-k} \cdot \chi \cdot \sum_I \prod_{i \notin I} \ell_i. \quad \square
\end{aligned}$$

In the next step we generalize the statement of Lemma 41 to sums of rectangular functions.

Corollary 42. *Let N_1, \dots, N_n be positive integers, let $\phi_{i,k} \geq 0$ be a non-negative real, let $I_{i,k} \subseteq \mathbb{R}$ be an interval of length $\ell_{i,k}$, and let $f_{i,k}: \mathbb{R} \rightarrow \mathbb{R}$ be the function*

$$f_{i,k}(x) = \begin{cases} \phi_{i,k} & \text{if } x \in I_{i,k}, \\ 0 & \text{otherwise,} \end{cases}$$

where $i \in [n]$, $k \in [N_i]$. Furthermore, let $f_i: \mathbb{R} \rightarrow \mathbb{R}$ be the function $f_i = \sum_{k=1}^{N_i} f_{i,k}$, let $f: \mathbb{R}^n \rightarrow \mathbb{R}$ be the function $f(x_1, \dots, x_n) = \prod_{i=1}^n f_i(x_i)$, and let $A \in \{-\mathcal{K}, \dots, \mathcal{K}\}^{n \times n}$ be an invertible matrix. Then

$$\int_{y \in \mathbb{R}^{n-k}} \max_{z \in \mathbb{R}^k} f(A^{-1} \cdot (y, z)) dy \leq 2^k \cdot (n-k)! \cdot \mathcal{K}^{n-k} \cdot \sum_I \left(\left(\prod_{i \notin I} \sigma_i \right) \cdot \left(\prod_{i \in I} \chi_i \right) \right)$$

where $\sigma_i = \sum_{k=1}^{N_i} \phi_{i,k} \cdot \ell_{i,k}$ and $\chi_i = \sum_{k=1}^{N_i} \phi_{i,k}$ and where the first sum runs over all tuples $I = (i_1, \dots, i_k)$ for which $1 \leq i_1 < \dots < i_k \leq n$.

Proof. For indices $k_i \in [N_i]$ let $f_{k_1, \dots, k_n}(x_1, \dots, x_n) = \prod_{i=1}^n f_{i, k_i}(x_i)$. This function is of the form assumed in Lemma 41 and takes only values 0 and $\chi_{k_1, \dots, k_n} = \prod_{i=1}^n \phi_{i, k_i}$. We can write function f as

$$\begin{aligned}
f(x_1, \dots, x_n) &= \prod_{i=1}^n f_i(x_i) = \prod_{i=1}^n \sum_{k_i=1}^{N_i} f_{i, k_i}(x_i) = \sum_{k_1=1}^{N_1} \dots \sum_{k_n=1}^{N_n} \prod_{i=1}^n f_{i, k_i}(x_i) \\
&= \sum_{k_1=1}^{N_1} \dots \sum_{k_n=1}^{N_n} f_{k_1, \dots, k_n}(x_1, \dots, x_n).
\end{aligned}$$

For the sake of simplicity we write \sum_{k_i} instead of $\sum_{k_i=1}^{N_i}$ and $\sum_{k_i: i \in (i_1, \dots, i_\ell)}$ instead of $\sum_{k_{i_1}} \dots \sum_{k_{i_\ell}}$. We can bound the integral as follows:

$$\begin{aligned}
\int_{y \in \mathbb{R}^{n-k}} \max_{z \in \mathbb{R}^k} f(A^{-1} \cdot (y, z)) dy &= \int_{y \in \mathbb{R}^{n-k}} \max_{z \in \mathbb{R}^k} \sum_{k_i: i \in [n]} f_{k_1, \dots, k_n}(A^{-1} \cdot (y, z)) dy \\
&\leq \int_{y \in \mathbb{R}^{n-k}} \sum_{k_i: i \in [n]} \max_{z \in \mathbb{R}^k} f_{k_1, \dots, k_n}(A^{-1} \cdot (y, z)) dy \\
&= \sum_{k_i: i \in [n]} \int_{y \in \mathbb{R}^{n-k}} \max_{z \in \mathbb{R}^k} f_{k_1, \dots, k_n}(A^{-1} \cdot (y, z)) dy \\
&\leq \sum_{k_i: i \in [n]} \left(2^k \cdot (n-k)! \cdot \mathcal{K}^{n-k} \cdot \chi_{k_1, \dots, k_n} \cdot \sum_I \prod_{i \notin I} \ell_{i, k_i} \right) \\
&= 2^k \cdot (n-k)! \cdot \mathcal{K}^{n-k} \cdot \sum_{k_i: i \in [n]} \left(\prod_{i \in [n]} \phi_{i, k_i} \cdot \sum_I \prod_{i \notin I} \ell_{i, k_i} \right) \\
&= 2^k \cdot (n-k)! \cdot \mathcal{K}^{n-k} \cdot \sum_I \sum_{k_i: i \in [n]} \left(\prod_{i \in [n]} \phi_{i, k_i} \cdot \prod_{i \notin I} \ell_{i, k_i} \right),
\end{aligned}$$

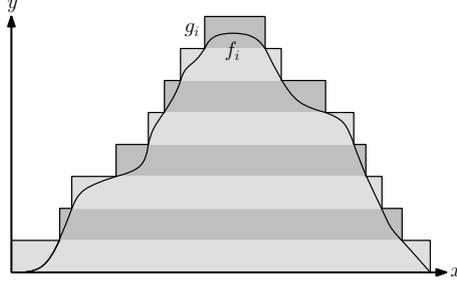


Figure 3: Area of a quasi-concave function covered by a “stack” of rectangles with approximately the same area

where the second inequality is due to Lemma 41. Now,

$$\begin{aligned}
\sum_{k_i: i \in [n]} \left(\prod_{i \in [n]} \phi_{i,k_i} \cdot \prod_{i \notin I} \ell_{i,k_i} \right) &= \sum_{k_i: i \in [n]} \left(\prod_{i \in I} \phi_{i,k_i} \cdot \prod_{i \notin I} (\phi_{i,k_i} \cdot \ell_{i,k_i}) \right) \\
&= \left(\sum_{k_i: i \in I} \prod \phi_{i,k_i} \right) \cdot \left(\sum_{k_i: i \notin I} \prod (\phi_{i,k_i} \cdot \ell_{i,k_i}) \right) \\
&= \left(\prod_{i \in I} \sum_{k_i} \phi_{i,k_i} \right) \cdot \left(\prod_{i \notin I} \sum_{k_i} (\phi_{i,k_i} \cdot \ell_{i,k_i}) \right) \\
&= \left(\prod_{i \in I} \chi_i \right) \cdot \left(\prod_{i \notin I} \sigma_i \right),
\end{aligned}$$

which completes the proof of Corollary 42. \square

To finish the proof of Theorem 40 we round the probability densities f_i as follows: For an arbitrarily small positive real δ let $g_i := \lceil f_i/\delta \rceil \cdot \delta$, i.e., we round f_i up to the next integral multiple of δ . As the densities f_i are quasiconcave, there is a decomposition of g_i such that $g_i = \sum_{k=1}^{N_i} f_{i,k}$ where

$$f_{i,k} = \begin{cases} \phi_{i,k} & : x \in I_{i,k}, \\ 0 & : \text{otherwise,} \end{cases} \quad \text{and} \quad \chi_i := \sum_{k=1}^{N_i} \phi_{i,k} = \max_{x \in [-1,1]} g_i(x),$$

where $I_{i,k}$ are intervals of length $\ell_{i,k}$ and $\phi_{i,k}$ are positive reals. The second property is the interesting one and stems from the quasiconcaveness of f_i . Informally speaking the two-dimensional shape bounded by the horizontal axis and the graph of g_i is a stack of rectangles aligned with axes (see Figure 3). Therefore, the sum χ_i of the rectangles' heights which appears in the formula of Corollary 42 is approximately ϕ . Without the quasiconcaveness χ_i might be unbounded.

Applying Corollary 42, we obtain

$$\begin{aligned}
\int_{y \in \mathbb{R}^{n-k}} \max_{z \in \mathbb{R}^k} f_X(A^{-1} \cdot (y, z)) dy &= \int_{y \in \mathbb{R}^{n-k}} \max_{z \in \mathbb{R}^k} \prod_{i=1}^n f_i((A^{-1} \cdot (y, z))_i) dy \\
&\leq \int_{y \in \mathbb{R}^{n-k}} \max_{z \in \mathbb{R}^k} \prod_{i=1}^n g_i((A^{-1} \cdot (y, z))_i) dy \\
&\leq 2^k \cdot (n-k)! \cdot \mathcal{K}^{n-k} \cdot \sum_I \left(\prod_{i \notin I} \sum_{k_i=1}^{N_i} (\phi_{i,k_i} \cdot \ell_{i,k_i}) \right) \cdot \left(\prod_{i \in I} \chi_i \right)
\end{aligned}$$

$$= 2^k \cdot (n-k)! \cdot \mathcal{K}^{n-k} \cdot \sum_I \left(\prod_{i \notin I} \int_{[-1,1]} g_i dx \right) \cdot \left(\prod_{i \in I} \chi_i \right).$$

Since $0 \leq \int_{[-1,1]} g_i dx \leq \int_{[-1,1]} (f_i + \delta) dx = 1 + 2\delta$ and $0 \leq \chi_i \leq \sup_{x \in [-1,1]} f_i(x) + \delta \leq \phi + \delta$, this implies

$$\begin{aligned} \int_{y \in \mathbb{R}^{n-k}} \max_{z \in \mathbb{R}^k} f_X(A^{-1} \cdot (y, z)) dy &\leq 2^k \cdot (n-k)! \cdot \mathcal{K}^{n-k} \cdot \sum_I \left(\prod_{i \notin I} (1+2\delta) \right) \cdot \left(\prod_{i \in I} (\phi + \delta) \right) \\ &= 2^k \cdot (n-k)! \cdot \mathcal{K}^{n-k} \cdot \sum_I (1+2\delta)^{n-k} \cdot (\phi + \delta)^k \\ &= 2^k \cdot (n-k)! \cdot \mathcal{K}^{n-k} \cdot \binom{n}{k} \cdot (1+2\delta)^{n-k} \cdot (\phi + \delta)^k \\ &\leq 2^k \cdot n^{n-k} \cdot \mathcal{K}^{n-k} \cdot (1+2\delta)^{n-k} \cdot (\phi + \delta)^k. \end{aligned}$$

As this bound is true for arbitrarily small reals $\delta > 0$, we obtain the desired bound of

$$2^k (n\mathcal{K})^{n-k} \phi^k = 2^k (m\mathcal{K})^{m-k} \phi^k. \quad \square$$

8 Conclusions and Open Problems

With the techniques developed in this article we settled two questions posed by Moitra and O'Donnell [14]: For quasiconcave densities we showed that the exponent of ϕ in the bound for the smoothed number of Pareto-optimal solutions is exactly d . Moreover, we significantly improved on the previously best known bound for higher moments of the smoothed number of Pareto-optima by Röglin and Teng [18].

Maybe even more interesting are our results for the model of zero-preserving perturbations suggested by Spielman and Teng [21] and Beier and Vöcking [4]. For this model we proved the first non-trivial bound on the smoothed number of Pareto-optimal solutions. We showed that this result can be used to analyze multiobjective optimization problems with polynomial and even more general objective functions. Furthermore, our result implies that the smoothed running time of the algorithm proposed by Berger et al. [5] to compute a path trade in a routing network is polynomially bounded for every constant number of autonomous systems. We believe that there are many more such applications of our result in the area of multiobjective optimization.

There are several interesting open questions. First of all it would be interesting to find asymptotically tight bounds for the smoothed number of Pareto-optimal solutions. There is still a gap between our upper bound of $O(n^{2d}\phi^d)$ for quasiconcave ϕ -smooth instances and the best lower bound of $\Omega(n^{d-1.5}\phi^d)$ [6]. Only for the case $d = 1$ we can show that the upper bound is tight [6].

Especially for zero-preserving perturbations there is still a lot of work to do. We conjecture that our techniques can be extended to also bound higher moments of the smoothed number of Pareto-optima for ϕ -smooth instances with zero-preserving perturbations. However, we feel that even our bound for the first moment is too pessimistic as we do not have a lower bound showing that setting coefficients to 0 can lead to larger Pareto sets. It would be very interesting to either prove a lower bound that shows that zero-preserving perturbations can lead to larger Pareto-sets than non-zero-preserving perturbations or to prove a better upper bound for zero-preserving perturbations.

References

- [1] René Beier. *Probabilistic Analysis of Discrete Optimization Problems*. PhD thesis, Universität des Saarlandes, 2004.
- [2] René Beier, Heiko Röglin, and Berthold Vöcking. The smoothed number of Pareto optimal solutions in bicriteria integer optimization. In *Proceedings of the 12th International Conference on Integer Programming and Combinatorial Optimization (IPCO)*, pages 53–67, 2007.
- [3] René Beier and Berthold Vöcking. Random knapsack in expected polynomial time. *Journal of Computer and System Sciences*, 69(3):306–329, 2004.
- [4] René Beier and Berthold Vöcking. Typical properties of winners and losers in discrete optimization. *SIAM Journal on Computing*, 35(4):855–881, 2006.
- [5] André Berger, Heiko Röglin, and Ruben van der Zwaan. Path trading: Fast algorithms, smoothed analysis, and hardness results. In *Proceedings of the 10th International Symposium on Experimental Algorithms (SEA)*, pages 43–53, 2011.
- [6] Tobias Brunsch, Navin Goyal, Luis Rademacher, and Heiko Röglin. Lower bounds for the average and smoothed number of pareto-optima. *Theory of Computing*, 2014. to appear.
- [7] Tobias Brunsch and Heiko Röglin. Improved smoothed analysis of multiobjective optimization. In *Proceedings of the 44th Annual ACM Symposium on Theory of Computing (STOC)*, pages 407–426, 2012.
- [8] H. William Corley and I. Douglas Moon. Shortest paths in networks with vector weights. *Journal of Optimization Theory and Application*, 46(1):79–86, 1985.
- [9] Matthias Ehrgott. Integer solutions of multicriteria network flow problems. *Investigacao Operacional*, 19:229–243, 1999.
- [10] Matthias Ehrgott. *Multicriteria Optimization*. Springer, 2005.
- [11] Matthias Ehrgott and Xavier Gandibleux. Multiobjective combinatorial optimization. In Matthias Ehrgott and Xavier Gandibleux, editors, *Multiple Criteria Optimization – State of the Art Annotated Bibliographic Surveys*, pages 369–444. Kluwer Academic Publishers, 2002.
- [12] Pierre Hansen. Bicriterion path problems. In *Multiple Criteria Decision Making: Theory and Applications*, volume 177 of *Lecture Notes in Economics and Mathematical Systems*, pages 109–127, 1980.
- [13] Kathrin Klamroth and Margaret M. Wiecek. Dynamic programming approaches to the multiple criteria knapsack problem. *Naval Research Logistics*, 47(1):57–76, 2000.
- [14] Ankur Moitra and Ryan O’Donnell. Pareto optimal solutions for smoothed analysts. *SIAM Journal on Computing*, 41(5):1266–1284, 2012.
- [15] Matthias Müller-Hannemann and Karsten Weihe. Pareto shortest paths is often feasible in practice. In *Proceedings of the 5th International Workshop on Algorithm Engineering (WAE)*, pages 185–198, 2001.
- [16] Adli Mustafa and Mark Goh. Finding integer efficient solutions for bicriteria and tricriteria network flow problems using dinas. *Computers & Operations Research*, 25(2):139–157, 1998.
- [17] George L. Nemhauser and Zev Ullmann. Discrete dynamic programming and capital allocation. *Management Science*, 15(9):494–505, 1969.

- [18] Heiko Röglin and Shang-Hua Teng. Smoothed analysis of multiobjective optimization. In *Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 681–690, 2009.
- [19] Arvind Sankar, Daniel A. Spielman, and Shang-Hua Teng. Smoothed analysis of the condition numbers and growth factors of matrices. *SIAM Journal on Matrix Analysis Applications*, 28(2):446–476, 2006.
- [20] Anders J. V. Skriver and Kim Allan Andersen. A label correcting approach for solving bicriterion shortest-path problems. *Computers & Operations Research*, 27(6):507–524, 2000.
- [21] Daniel A. Spielman and Shang-Hua Teng. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. *Journal of the ACM*, 51(3):385–463, 2004.
- [22] Daniel A. Spielman and Shang-Hua Teng. Smoothed analysis: an attempt to explain the behavior of algorithms in practice. *Communications of the ACM*, 52(10):76–84, 2009.