Pareto Optimal Solutions for Smoothed Analysts

Ankur Moitra* Ryan O'Donnell†

November 11, 2010

Abstract

Consider an optimization problem with n binary variables and $d+1$ linear objective functions. Each valid solution $x \in \{0,1\}^n$ gives rise to an objective vector in $\mathbb{R}^{d+1}$, and one often wants to enumerate the Pareto optima among them. In the worst case there may be exponentially many Pareto optima; however, it was recently shown that in (a generalization of) the smoothed analysis framework, the expected number is polynomial in n. Unfortunately, the bound obtained had a rather bad dependence on d; roughly n^{d^d}. In this paper we show a significantly improved bound of n^{2d}.

Our proof is based on analyzing two algorithms. The first algorithm, on input a Pareto optimal x, outputs a "testimony" containing clues about x's objective vector, x's coordinates, and the region of space B in which x's objective vector lies. The second algorithm can be regarded as a *speculative* execution of the first — it can uniquely reconstruct x from the testimony's clues and just *some* of the probability space's outcomes. The remainder of the probability space's outcomes are just enough to bound the probability that x's objective vector falls into the region B.

*Massachusetts Institute of Technology, moitra@mit.edu. Part of this research done while at Microsoft Research New England. Supported in part by a Fannie and John Hertz Foundation Fellowship.

†Carnegie Mellon University, odonnell@cs.cmu.edu. Part of this research done while visiting Microsoft Research New England; part done while at the Institute for Advanced Study. Supported by NSF grants CCF-0747250 and CCF-0915893, BSF grant 2008477, and Sloan and Okawa fellowships.

1 Introduction

We study the expected number of Pareto optimal solutions in multiobjective binary optimization problems within the framework of *smoothed analysis*.

1.1 Multiobjective optimization and Pareto optima

In a typical decision-making problem there are multiple criteria used in judging the quality of a solution. For example, in choosing a driving route between two points one might want to minimize distance, tolls, number of turns, and expected traffic; in choosing a vacation hotel one might want to minimize price and distance to the beach, while maximizing quality. In such cases there is rarely a single solution which is best on all criteria simultaneously. The most popular way to handle the tradeoff is to determine the set of all *Pareto optimal* solutions, meaning those solutions which are not dominated in all measures of quality by some other solution. This idea, originating in microeconomics, has been very extensively studied in computer science, especially in operations research [Ehr05], algorithmic theory [PY02], artificial intelligence [Deb01], and database theory (under the name "skyline queries") [BKS01].

Even if one is not interested in Pareto optima per se, many algorithms and heuristics for solving optimization problems enumerate Pareto optimal solutions as an intermediary step. For example, the Nemhauser–Ullmann algorithm [NU69] for exactly solving the Knapsack problem works by iteratively computing the Pareto optimal $\langle \text{value}, \text{weight} \rangle$ pairs achievable by the first i items, for $i = 1 \dots n$. Beier and Vöcking [BV04] observed that this algorithm runs in time $O(nB)$, where B is an upper bound on the number of Pareto optima in each stage. As we describe below, this allowed them to give the first polynomial-time algorithm for an NP-hard optimization problem in the smoothed analysis framework, by performing smoothed analysis on the number of Pareto optimal solutions.

Unfortunately, even in the simplest case multiobjective optimization — two linear objective functions — the number of Pareto optimal solutions may be exponentially large in the number of decision variables. There have been two main approaches to dealing with this exponential complexity. The first, popularized by Papadimitriou and Yannakakis [PY02], involves computing "ϵ-approximate Pareto sets". In many cases, polynomial-size ϵ-approximate Pareto sets can be computed efficiently; see the thesis of Diakonikolas [Dia10] for references. The second approach, pioneered by Beier and Vöcking [BV04], involves studying multiobjective optimization in the smoothed analysis framework.

1.2 Smoothed analysis for discrete optimization

Smoothed analysis was introduced in an influential work of Spielman and Teng [ST04], as a framework intermediate between worst-case and average-case analysis. Here the idea is to think of real numbers in the input as being defined based on imprecise measurements; specifically, they are first fixed adversarially in $[-1, 1]$, say, and then subjected to Gaussian noise with some small standard deviation σ. In this framework, Spielman and Teng showed that a certain version of the simplex algorithm for linear programming runs in $\text{poly}(n, 1/\sigma)$ expected time.

A notable work of Beier and Vöcking [BV04] from 2003 showed that the NP-hard 0/1-Knapsack problem can be solved in polynomial time in the smoothed analysis framework. (Previously, there had been a long line of work on average-case analysis of 0/1-Knapsack: see, e.g., [DF89, GMS84, Lue98].) Furthermore, they showed this holds even in a much more general model of smoothed analysis. In one version of their model, each item's profit P_i and weight W_i is an independent

random variable with values in $[0,1]$; the only restriction is that the probability density function (pdf) of each P_i and W_i is upper-bounded by the parameter ϕ. We call this model "ϕ-semirandom". It is easy to see that as ϕ is increased, the framework goes from (a very general version of) average-case analysis to worst-case analysis. For example, given a small number σ, if we take $\phi = 1/\sigma$ then the profits P_i could be of the form $p_i + U_i$, where $p_i \in [\sigma, 1-\sigma]$ is an adversarially chosen number and U_i is uniformly random on $[-\sigma, \sigma]$. (The original case of Gaussian noise does not quite fit in this framework, but is easily handled with a small amount of additional work.)

1.3 Previous work

Beier and Vöcking showed that in this ϕ-semirandom model, the expected number of Pareto optimal knapsacks is $O(\phi n^4)$; from this they immediately deduced that the Nemhauser–Ullmann algorithm runs in expected $O(\phi n^5)$ time. In fact, Beier and Vöcking showed that the same is true even if the weights are *adversarially* specified, and only the profits are chosen randomly (independently, from ϕ-bounded distributions). In this case of adversarially weights, they also showed an $\Omega(n^2)$ lower bound for the expected number of Pareto optima, even for uniformly distributed profits (i.e., $\phi = 1$).

In his thesis, Beier [Bei04] extended this analysis to general 2-objective binary optimization problems. Specifically, he showed that given an arbitrary set of "solutions" $\mathcal{S} \subseteq \{0,1\}^n$ and arbitrary 2nd objective values $\mathrm{Obj}^2(x)$ for each $x \in \mathcal{S}$, if the 1st objective is linear and ϕ-semirandom, then the expected number of Pareto optima is still $O(\phi n^4)$. Later work of Beier, Röglin, and Vöcking [BRV07] improved this bound to $O(\phi n^2)$ (which is tight for constant ϕ) and also extended to the case of integer-valued decision variables.

These works only handled the case of 2 objectives. Recently, Röglin and Teng [RT09] extended the analysis to the case in which there are $d+1$ objective functions, d of which are linear and ϕ-semirandom, and one of which is completely arbitrary. Their bound on the expected number of Pareto optima is polynomial in n and ϕ for constant d, and they were also able to polynomially bound higher moments. Unfortunately, their result is probably of theoretical interest only, as the dependence on d is rather bad. E.g., for $d = 3$ their upper bound on the expected number of Pareto optima is roughly n^{97} assuming $n \geq 2^{453787938}$ (and is much worse than n^{97} for smaller n). In general their bound is roughly $(\sqrt{\phi} \cdot n)^{f(d)}$ for $f(d) = 2^{d-1}(d+1)!$, once $n \geq \exp(\exp(d^2 \log d))$. Röglin and Teng concluded their work by asking whether the exponent $f(d)$ on n could be reduced from $d^{\Theta(d)}$ to $\mathrm{poly}(d)$; this was later recognized as an important open problem [Ten10]. Here, we resolve this question.

Very closely related to the research we have just described is a sequence of works [BV06, ANRV07, RV07, RT09], starting with Beier and Vöcking and culminating with Röglin and Teng, showing that binary optimization problems are solvable in expected polynomial time in the smoothed analysis framework if and only if they are solvable in randomized pseudopolynomial time in the worst case.

1.4 Our contribution

In this work we give an affirmative answer to the open problem of Röglin and Teng, reducing their bound from roughly $n^{2^{d-1}(d+1)!}$ to n^{2d}. Thus the exponent on n can in fact be made *linear* in d. In particular, we prove that the expected number of Pareto optimal solutions in the model described above is at most $2 \cdot (4\phi d)^{d(d+1)/2} \cdot n^{2d}$. It is interesting to compare our result with what is known about Pareto optima when 2^n points are chosen independently and uniformly in $[-1,1]^{d+1}$. In this scenario, old results [BKST78, Dev80, Buc89] show that the expected number of Pareto optima is

$\Theta(n)^d$ for each constant d. Our bound is within a square of this quantity, despite the significant dependencies in the model. We also note that this square is necessary at least for $d = 1$, due to the $\Omega(n^2)$ lower bound of Beier and Vöcking [BV04].

Usually, in smoothed analysis we are interested in demonstrating that a certain algorithm runs quickly or that a certain approximation algorithm returns a near-optimal solution. In such cases, one often defines an event – some property of the data that ensures an algorithm runs quickly or an approximation algorithm works well. This is true in the context of previous literature on the number of Pareto optimal solutions as well — indeed, the works of [BV06, RT09] are based on notions of *winner gap* and *loser gap* which can be interpreted as a discrete analogue to condition number.

Our approach turns this around: We give a deterministic algorithm, which on input a Pareto optimal x, runs on the data and produces an event – in the form of a "testimony" containing clues about x's objective vector, x's coordinates, and the region of space B in which x's objective vector lies. Our family of events is rather complicated, but is defined implicitly based on a simple algorithm.

We then give a second algorithm which can be regarded as a *speculative* execution of the first — it can uniquely reconstruct x from the testimony's clues and just *some* of the probability space's outcomes. The remainder of the probability space's outcomes are just enough to bound the probability that x's objective vector falls into the region B. So we are able to bound the probability that any particular "testimony" is output by the first algorithm, and consequently we are able to give an upper bound on the expected number of Pareto optimal solutions.

2 Our result and approach

In this section we will describe the problem formally, state our Main Theorem, and then briefly describe our approach. The remainder of the paper is devoted to the proof of the Main Theorem.

2.1 Problem definitions

Our setting captures the broad class of multiobjective binary optimization problems in which all (but one) of the objective functions are linear. We fix once and for all an arbitrary set $\mathcal{S} \subseteq \{0,1\}^n$ of *solutions*. These might encode knapsacks, the sets of edges forming a spanning tree in a given graph, or even the sets of edges forming a Hamiltonian cycle.

Matrix notation. We think of solutions in $\mathcal{S} \subseteq \{0,1\}^n$ as column vectors. For a matrix (or vector) A, we will write A^i for the i'th row of A and write A_j for the j'th column of A; thus A^i_j is the (i,j) entry of A. For $i < k$ we will also write $A^{i..k}$ for the submatrix of A consisting of rows i through k. Given matrices A and B of the same size we write $A \circ B$ for their Hadamard product, i.e., their entry-wise product. Thus $(A \circ B)^i_j = A^i_j B^i_j$.

Values and objectives. Associated to each solution $x \in \mathcal{S}$ are $d+1$ *objectives*; we encode them with a column vector $\mathrm{Obj}(x) \in \mathbb{R}^{d+1}$. The first d objectives are assumed to be linear and are chosen in a "ϕ-semirandom" fashion. More specifically, there is a $d \times n$ matrix $\boldsymbol{V}$ of random variables called *values*. (We typically write random variables in boldface.) We assume that each entry of $\boldsymbol{V}$ is an independent, continuous random variable with support on $[-1,1]$ and pdf bounded by the parameter ϕ. The first d objectives of solution x are defined by the equation $\mathrm{Obj}^{1..d}(x) = \boldsymbol{V}x$. (Recall that $x \in \{0,1\}^n$ is thought of as a column vector.) The $(d+1)$'th objectives of the solutions

are neither linear nor random. We assume merely that they are fixed, distinct real numbers, chosen in advance of $\boldsymbol{V}$. (Indeed, their magnitudes are not important for us, only their relative ordering.) We will significantly abuse notation by writing $\boldsymbol{V}^{d+1}x$ in place of $\mathrm{Obj}^{d+1}(x)$. In this way, $\mathrm{Obj}^i(x) = \boldsymbol{V}^i x$ holds for each $i \in [d+1]$.

Pareto optima. Without loss of generality, we think of higher objectives as preferable. Accordingly, given (column) vectors $p, q \in \mathbb{R}^{d+1}$ we say that p *dominates* q if $p \geq q$. Here "$\geq$" is to be interpreted entry-wise when applied to vectors; i.e., p dominates q if $p^i \geq q^i$ for all $i \in [d+1]$. We will also sometimes use the notion of *t-domination* for $t < d+1$; we say that p t-dominates q if $p^{1..t} \geq q^{1..t}$. Given a set of points $\mathcal{P} \subset \mathbb{R}^{d+1}$ we say that $p \in \mathcal{P}$ is *Pareto optimal (within $\mathcal{P}$)* if p is not dominated by any other point $q \in \mathcal{P}$; i.e., for all $q \in \mathcal{P} \setminus \{p\}$, there exists $i \in [d+1]$ with $p^i > q^i$. Of course, we will be interested in applying this concept to the objectives of the solutions in $\mathcal{S}$. Given $\boldsymbol{V}$, we consider $\mathcal{P} = \{\mathrm{Obj}(z) : z \in \mathcal{S}\} \subset [-n, n]^d \times \mathbb{R}$. We then say that the *solution* $x \in \mathcal{S}$ is Pareto optimal if $\mathrm{Obj}(x)$ is Pareto optimal within $\mathcal{P}$. Finally, given $\boldsymbol{V}$, we define $\mathbf{PO} \subseteq \mathcal{S}$ to be the set of all Pareto optimal solutions.

2.2 Our result

We can now state our Main Theorem:

Main Theorem. $\displaystyle \mathop{\mathbf{E}}_{\boldsymbol{V}}\Big[\big|\mathbf{PO}\big|\Big] \leq 2 \cdot (4\phi d)^{d(d+1)/2} \cdot n^{2d}.$

2.3 Our approach

To prove the Main Theorem we use a probabilistic argument which has a rather unusual form. Unfortunately, it is also fairly intricate. In this section we will try to convey some of the ideas of the argument while hiding a number of complicating details.

Our proof can be seen as a d-dimensional generalization of the Beier–Röglin–Vöcking $O(\phi n^2)$ upper bound for the $d = 1$ case (which we will later sketch). The reader is advised to keep the cases $d = 1, 2$ in mind for visualization purposes. Recall that the solutions $x \in \mathcal{S}$ have d semirandom linear objectives but their $(d+1)$'th objectives are fixed in advance arbitrarily. Once the values $\boldsymbol{V}$ are drawn and the objectives $\mathrm{Obj}^{1..d}(x) \in [-n, n]^d$ thus determined, one can think of identifying the Pareto optima among $\mathcal{S}$ via a "sweep" along the $(d+1)$'th dimension. This means proceeding through the solutions $x \in \mathcal{S}$ in decreasing order of $\mathrm{Obj}^{d+1}(x)$ and considering the "point" $\mathrm{Obj}^{1..d}(x) = \boldsymbol{V}x \in [-n, n]^d$; the set of points which are not d-dominated by any previously seen point correspond exactly to the set of Pareto optimal solutions.

Boxes and density. An oversimplification of our proof is to think of it as showing that the "probability density" of Pareto optimal points in $[-n, n]^d$ is not too high; roughly $O(n^d)$. In aid of making this formal, we fix once and for all a real number $\epsilon > 0$ which should be thought of as extremely small, $\epsilon \ll 1/(\phi d 2^{2n})$. Additionally, we assume that $1/\epsilon$ is an integer. We then introduce the following definition:

Definition 2.1. For a point $b \in (\epsilon\mathbb{Z})^d$, we define the *$d$-box based at point b* to be $b + [0, \epsilon)^d$. Note that the set of all d-boxes partitions $[-n, n)^d$ and indeed all of $\mathbb{R}^d$. More generally, for $t \in [d]$ and $b \in (\epsilon\mathbb{Z})^j$, we define the *$t$-box based at point b* to be $(b + [0, \epsilon)^d) \times \mathbb{R}^{d-t}$. The set of all t-boxes also partitions $\mathbb{R}^d$.

Since ϵ is so small, the probability that there will be two different points $\boldsymbol{V}x$ and $\boldsymbol{V}x'$ in the same d-box is negligible. Thus if B denotes an arbitrary d-box, we can upper-bound the number of Pareto optima by $(2n/\epsilon)^d$ times the probability that there is a Pareto optimum $x \in \mathcal{S}$ with $\mathrm{Obj}^{1..d}(x)$ in B. Our goal is to bound this probability by roughly $O(n^d)\epsilon^d$.

In particular, we must make sure to keep the probability roughly comparable to ϵ^d. A crucial aspect of our proof is that we can bound $\mathbf{Pr}[\boldsymbol{V}x \in B]$ by $(\phi\epsilon)^d$ for any $x \neq \vec{0}$ *while only using a small part of the probability space.* Specifically, suppose we select $j \in [n]$ such that $x^j \neq 0$, and then imagine drawing all entries of $\boldsymbol{V}$ except for the j'th column $\boldsymbol{V}_j$. Then the final position of the point $\boldsymbol{V}x$ is of the form $(p^1 + \boldsymbol{V}^1_j, \dots, p^n + \boldsymbol{V}^n_j)$, where the p^i's are constants. This point will lie in the box B only if each value $\boldsymbol{V}^i_j$ falls into a certain fixed interval of width ϵ. As the random variables $\boldsymbol{V}^i_j$ are independent and have pdf's bounded by ϕ, the probability that all $\boldsymbol{V}^i_j$'s fall into the required intervals is at most $(\phi\epsilon)^d$. Note that this argument works for any possible outcome of the $d(n-1)$ values outside of $\boldsymbol{V}_j$.

Uniqueness. Unfortunately we cannot simply take this observation and union-bound over all potential Pareto optima x, since this would lose a factor of $|\mathcal{S}|$. We would be in much better shape if, after all values except for $\boldsymbol{V}_j$ were drawn, there were very few solutions x — or even just a unique solution x — for which the event

$$\mathbf{T}_x = \text{“}x \text{ is Pareto optimal with } \boldsymbol{V}x \in B\text{”}$$

had a chance of occurring. Here by “have a chance of occurring”, we mean $\mathbf{Pr}_{\boldsymbol{V}_j}[\mathbf{T}_x] > 0$. In the simplest case of $d = 1$, Beier, Röglin, and Vöcking [BRV07] essentially show that essentially holds if one adds some extra conditions to the event $\mathbf{T}_x$. We now sketch a reinterpretation of their argument.

The Beier–Röglin–Vöcking argument. Note that since $d = 1$ for this sketch, the values matrix $\boldsymbol{V}$ is just a random (row) vector. For each $j \in [n]$ and 1-box (interval) B, let us define the significantly more complicated event

$$\mathbf{T}_{x,j,B} = \text{“}x^j = 1,\ \boldsymbol{V}x \in B,\ x \text{ is Pareto optimal, and the ‘next’ Pareto optimum } y \text{ has } y^j = 0\text{”}.$$

Here ‘next’ refers to the “sweep along the 2nd coordinate”; i.e., y is the solution z with maximal $\mathrm{Obj}^2(z)$ among $\{z \in \mathcal{S} : \boldsymbol{V}z > \boldsymbol{V}x\}$. The Beier–Röglin–Vöcking argument takes a union bound over all $j \in [n]$ in addition to over all B. The key to their argument is the following “uniqueness” claim: for any draw of the values other than $\boldsymbol{V}_j$, there is a *unique* $x \in \mathcal{S}$ for which the event $\mathbf{T}_{x,j,B}$ has a chance of occurring. Given this claim, the proof is almost complete. For that unique x the event $\mathbf{T}_{x,j,B}$ still has at most a $\phi\epsilon$ chance of occurring, since x^j must be 1 and the value $\boldsymbol{V}_j$ is still independent and undrawn. Union-bounding over all j and B, one concludes that the expected value of

$$\#\{\text{Pareto optimal } x : \text{the ‘next’ Pareto optimum } y \text{ has } y^j \neq 1 = x^j \text{ for some } j\}$$

is at most $n \cdot (2n/\epsilon) \cdot (\phi\epsilon) = 2\phi n^2$. This *almost* counts the total number of Pareto optima. Certainly for each Pareto optimum x, there is *some* coordinate j such that the ‘next’ Pareto optimum y has $y^j \neq a = x^j$; it's just that this bit a might be 0 rather than 1. The Beier–Röglin–Vöcking is concluded (essentially) by union-bounding over $a \in \{0, 1\}$ as well. (It may seem crucial that x^j was 1 and not 0 when we observed that $\mathbf{Pr}_{\boldsymbol{V}_j}[\boldsymbol{V}x \in B] \leq \phi\epsilon$. This difficulty is overcome with an additional trick, changing the condition $\boldsymbol{V}x \in B$ in $\mathbf{T}_{x,j,B}$ to the condition $\boldsymbol{V}x - \boldsymbol{V}_j\overline{a}$ in $\mathbf{T}_{x,j,a,B}$.)

The Röglin–Teng argument. How can we generalize this argument to d dimensions? Röglin and Teng roughly take the following approach. First, they generalize the above argument to show that for $d = 1$, the expected c'th power of the number of Pareto optima is $(\phi n^2)^{c(1+o(1))}$. This gives them a concentration result for the number of Pareto optima. They then proceed by induction on the dimension d. In reducing from dimension d to $d-1$ there are two difficulties. First, instead of having a unique x to worry about as in the Beier–Röglin–Vöcking, they need to worry about all solutions in a $(d-1)$-dimensional Pareto set. One expects this not to be too large, by induction; however, their argument needs a high-probability result. Hence they need to inductively bound higher powers of the number of Pareto optima. This induction leads to the rather bad dependence on d. A second difficulty they face comes from their use of conditioning to separate the d'th dimension from the first $d-1$; this introduces dependencies that they must work to control.

Our argument. We define a family of events $\mathbf{T}_{x,J,A,\mathcal{B}}$. These events are again of the form "x falls into a box related to $\mathcal{B}$ and certain other lower-dimensional conditions happen". We need to define these other conditions in an extremely careful way so that the following holds:

> *Based on J, there is a way to partition the draw of $\boldsymbol{V}$ into two parts called $M(J)\circ\boldsymbol{V}$ and $M(J)\circ\boldsymbol{V}$. Given the outcome of $M(J)\circ\boldsymbol{V}$, there is a* **unique** *$x \in \mathcal{S}$ for which $\mathbf{T}_{x,J,A,\mathcal{B}}$ can occur. Furthermore, the randomness remaining in $M(J)\circ\boldsymbol{V}$ is such that the probability of $\mathbf{T}_{x,J,A,\mathcal{B}}$ can* **still be bounded** *by an appropriately small quantity.*

We manage to identify the necessary conditions; however they are complicated enough that they cannot be described with just a sentence of text. Instead, we come to the first unusual aspect of our argument; the extra conditions are of the form "*a certain deterministic algorithm* `Witness`*, when run with input x and $\boldsymbol{V}$, produces a certain output testimony*". At this point the reader might think that such conditions have no chance to satisfy the boxed property above: in particular, since `Witness` depends on $\boldsymbol{V}$, how can knowing its output still leave the $M(J)\circ\boldsymbol{V}$ part of the probability space free? We overcome this problem with a second unusual idea. We introduce *another* deterministic algorithm called `Reconstruct`, which takes as input the output $\texttt{Witness}(x,\boldsymbol{V})$, along with the outcome of $M(J)\circ\boldsymbol{V}$. We show that using just this information, `Reconstruct` can recover the input x, *assuming that it is Pareto optimal*. This ability to reverse-engineer x gives us the needed "uniqueness" property, and the fact that `Reconstruct` does not need to know $M(J)\circ\boldsymbol{V}$ – and yet this amount of remaining randomness is still enough to bound the probability that x falls into certain boxes.

3 Outline of the proof

At this point we move from intuition to precise details. In this section we give the overall structure of our proof of the Main Theorem. By the end of this outline we will have reduced it to a number of lemmas, which are then proven in the appendices of the paper.

3.1 Testimonies

The first key ingredient in our proof is a deterministic map we call `Witness`, which takes as input a solution $x \in \mathcal{S}$ and a fixed matrix of values V, and outputs a "testimony" $(J, A, \mathcal{B})$:

$$\texttt{Witness}\colon (x, V) \mapsto (J, A, \mathcal{B}).$$

(The map $\texttt{Witness}$ also depends on the fixed quantities n, ϵ, $\mathcal{S}$, and the $(d+1)$'th objectives $\mathrm{Obj}^{d+1}(z)$.) We will actually only care about the behavior of $\texttt{Witness}(x, V)$ when the values V make x into a Pareto optimum, but it is clearest to define the mapping for every pair of x and V.

Regarding the testimony itself, roughly speaking J is a list of d coordinates in $[n]$, A is a "diagonalization matrix" consisting of d bits per coordinate in J, and $\mathcal{B}$ is a list of t-boxes, one for each $t \in [d]$. *Very* roughly speaking, the meaning of $\texttt{Witness}(x, V) = (J, A, \mathcal{B})$ is that the bits $\{x^j : j \in J\}$ agree with certain bits in A and that Vx falls into the boxes in $\mathcal{B}$ — or rather, that a slight translation of Vx based on A falls into these boxes. Precise details are given in Section 4, but they are not important for understanding the outline of the proof. On first reading, one should think of the number of possible testimonies as something roughly like $n^{2d}/\epsilon^{d(d+1)/2}$.

3.2 The OK event

We will also need to define a simple event based on the random draw of $\boldsymbol{V}$ which we call **OK**. In studying Pareto optima we prefer not to distinguish between domination and "strict" domination. Luckily we don't have to: since the entries of $\boldsymbol{V}$ are continuous random variables, the probability that $\boldsymbol{V}^i x = \boldsymbol{V}^i y$ for any $i \in [d]$ and distinct $x, y \in \mathcal{S}$ is 0. Our event **OK**, which we now formally define, slightly generalizes this:

Definition 3.1. $\mathbf{OK} = \mathbf{OK}(\boldsymbol{V})$ is defined to be the event that $|\boldsymbol{V}^i x - \boldsymbol{V}^i y| > \epsilon$ for all $i \in [d]$ and distinct $x, y \in \mathcal{S}$.

We require the following simple lemma:

OK Lemma. $\mathbf{Pr}[\neg\mathbf{OK}] \leq \phi d 2^{2n+1}\epsilon$.

Proof: For each fixed $i \in [d]$ and distinct $x, y \in \{0,1\}^n$, we show that $\mathbf{Pr}[|\boldsymbol{V}^i x - \boldsymbol{V}^i y| \leq \epsilon] \leq 2\phi\epsilon$; the result then follows by a union bound. Since x and y are distinct we may select $j \in [n]$ such that $x^j - y^j = 1$, after possibly exchanging x and y. Now imagine that the values $\{\boldsymbol{V}^i_k : k \neq j\}$ are drawn first; then the event $|\boldsymbol{V}^i x - \boldsymbol{V}^i y| \leq \epsilon$ becomes of the form $|c + \boldsymbol{V}^i_j| \leq \epsilon$ for some constant c. By independence, the random variable $\boldsymbol{V}^i_j$ still has pdf bounded by ϕ, so this event has probability at most $\phi \cdot 2\epsilon$, as desired. ■

3.3 Proof of the Main Theorem

We are now able to outline the proof of the Main Theorem.

$$\mathop{\mathbf{E}}_{\boldsymbol{V}}\Big[\big|\mathbf{PO}\big|\Big] = \mathop{\mathbf{E}}_{\boldsymbol{V}}\Big[\big|\mathbf{PO}\big| \cdot \mathbf{1}[\mathbf{OK}]\Big] + \mathop{\mathbf{E}}_{\boldsymbol{V}}\Big[\big|\mathbf{PO}\big| \cdot \mathbf{1}[\neg\mathbf{OK}]\Big]. \tag{1}$$

Regarding the second term in (1), naively we have

$$\mathop{\mathbf{E}}_{\boldsymbol{V}}\Big[\big|\mathbf{PO}\big| \cdot \mathbf{1}[\neg\mathbf{OK}]\Big] \leq \mathop{\mathbf{E}}_{\boldsymbol{V}}\Big[2^n \cdot \mathbf{1}[\neg\mathbf{OK}]\Big] = 2^n\,\mathbf{Pr}[\neg\mathbf{OK}] \leq \phi d 2^{3n+1}\epsilon, \tag{2}$$

using the OK Lemma. As for the first (main) term in (1), we break it up according to the possible testimonies:

$$\mathop{\mathbf{E}}_{\boldsymbol{V}}\Big[\big|\mathbf{PO}\big| \cdot \mathbf{1}[\mathbf{OK}]\Big] = \sum_{(J,A,\mathcal{B})} \mathop{\mathbf{E}}_{\boldsymbol{V}}\left[\sum_{x \in \mathcal{S}} \mathbf{1}[x \in \mathbf{PO}] \cdot \mathbf{1}[\texttt{Witness}(x, \boldsymbol{V}) = (J, A, \mathcal{B})] \cdot \mathbf{1}[\mathbf{OK}]\right]. \tag{3}$$

For a given draw of values $\boldsymbol{V}$, it is possible to show that *if* the event **OK** occurs, then the different $x \in \mathbf{PO}$ generate unique testimonies $(J, A, \mathcal{B})$. (This follows from the Testimony–Determines–PO Lemma in Section 4.) In other words, for a fixed testimony $(J, A, \mathcal{B})$, after $\boldsymbol{V}$ is drawn there can be at most one $x \in \mathcal{S}$ for which the event

$$\big(x \in \mathbf{PO}\big) \wedge \big(\texttt{Witness}(x, \boldsymbol{V}) = (J, A, \mathcal{B})\big) \wedge \mathbf{OK}$$

occurred. This shows that (3) is at most the number of possible testimonies. Unfortunately, that is not a helpful bound because the number of possible testimonies includes the huge factor $(1/\epsilon)^{d(d+1)/2}$.

We now come to the key idea in the proof. For each fixed testimony $(J, A, \mathcal{B})$, we split up the draw of $\boldsymbol{V}$ into two stages in a way that depends on J. In the first stage, "most" of the dn entries of $\boldsymbol{V}$ are drawn; we denote these entries by $\overline{M(J)} \circ \boldsymbol{V}$ for reasons to be explained later. In the second stage, the remaining "few" entries of $\boldsymbol{V}$ are drawn (independently, of course); we denote this second set of entries by $M(J) \circ \boldsymbol{V}$. On first reading, one should think of "few" as meaning $d(d+1)/2$. Now the key idea is that the uniqueness property described above holds *even after just drawing* $\overline{M(J)} \circ \boldsymbol{V}$:

Uniqueness Lemma. *Fix a testimony $(J, A, \mathcal{B})$ and fix the outcome of $\overline{M(J)} \circ \boldsymbol{V}$. Then there exists at most one $x \in \mathcal{S}$ such that the event*

$$\big(x \in \mathbf{PO}\big) \wedge \big(\texttt{Witness}(x, \boldsymbol{V}) = (J, A, \mathcal{B})\big) \wedge \mathbf{OK}$$

can occur. Here the event's randomness is just the draw of $M(J) \circ \boldsymbol{V}$.

Based on this idea, we write (3) as

$$\sum_{(J,A,\mathcal{B})} \mathop{\mathbf{E}}_{\overline{M(J)} \circ \boldsymbol{V}} \left[\sum_{x \in \mathcal{S}} \Pr_{M(J) \circ \boldsymbol{V}} \left[\big(x \in \mathbf{PO}\big) \wedge \big(\texttt{Witness}(x, \boldsymbol{V}) = (J, A, \mathcal{B})\big) \wedge \mathbf{OK} \right] \right].$$

The Uniqueness Lemma says that for each choice of $(J, A, \mathcal{B})$ and $\overline{M(J)} \circ \boldsymbol{V}$, at most one of the probabilities in the above expression can be nonzero. Hence we may upper-bound (3) by

$$\sum_{(J,A,\mathcal{B})} \mathop{\mathbf{E}}_{\overline{M(J)} \circ \boldsymbol{V}} \left[\max_{x \in \mathcal{S}} \Pr_{M(J) \circ \boldsymbol{V}} \left[\big(x \in \mathbf{PO}\big) \wedge \big(\texttt{Witness}(x, \boldsymbol{V}) = (J, A, \mathcal{B})\big) \wedge \mathbf{OK} \right] \right]. \tag{4}$$

We now complete the proof by showing that there is enough randomness left in $M(J) \circ \boldsymbol{V}$ so that for any $x \in \mathcal{S}$, even the probability of the event $\texttt{Witness}(x, \boldsymbol{V}) = (J, A, \mathcal{B})$ is small. We bound this probability in terms of a parameter called $\dim(\mathcal{B})$, which we define in Section 4 For now, it suffices to know that $\dim(\mathcal{B})$ is an integer between 0 and $d(d+1)/2$; on first reading, one should think of it as simply always being $d(d+1)/2$.

Boundedness Lemma. *For every fixed $(J, A, \mathcal{B})$, outcome of $\overline{M(J)} \circ \boldsymbol{V}$, and $x \in \mathcal{S}$, it holds that*

$$\Pr_{M(J) \circ \boldsymbol{V}} [\texttt{Witness}(x, \boldsymbol{V}) = (J, A, \mathcal{B})] \leq \phi^{\dim(\mathcal{B})} \epsilon^{\dim(\mathcal{B})}.$$

Using this in (4) we upper-bound (3) by

$$\sum_{(J,A,\mathcal{B})} \phi^{\dim(\mathcal{B})} \epsilon^{\dim(\mathcal{B})}. \tag{5}$$

As mentioned, on first reading one should think of $\dim(\mathcal{B})$ as always being $d(d+1)/2$ and one should think of the number of possible testimonies as being roughly $n^{2d}/\epsilon^{d(d+1)/2}$. Thus (5) is roughly $\phi^{d(d+1)/2} \cdot n^{2d}$, comparable to the quantity in the Main Theorem. We will eventually do a more precise but straightforward estimation to bound (5) (and hence (3)):

Counting Lemma. *For a fixed n and ϵ,*

$$\sum_{\substack{\text{possible testimonies} \\ (J,A,\mathcal{B})}} \phi^{\dim(\mathcal{B})}\epsilon^{\dim(\mathcal{B})} \leq 2 \cdot (4d\phi)^{d(d+1)/2} \cdot n^{2d}.$$

Substituting this bound on (3), as well as the bound (2), into (1) yields

$$\mathop{\mathbf{E}}_{\boldsymbol{V}}\Big[\big|\mathbf{PO}\big|\Big] \leq 2 \cdot (4d\phi)^{d(d+1)/2} \cdot n^{2d} + \phi d 2^{3n+1}\epsilon.$$

Since we can make ϵ arbitrarily small, the proof of the Main Theorem is complete.

4 Testimonies

In this section we describe the `Witness` algorithm, which assumes n, ϵ, $\mathcal{S}$, and the $(d+1)$'th objectives $\mathrm{Obj}^{d+1}(z)$ are fixed. The input to `Witness` is a solution $x \in \mathcal{S}$ and a $d \times n$ matrix of values V. The output is a "testimony", which is a triple $(J, A, \mathcal{B})$.

4.1 Components of a testimony

We now describe the components of a testimony.

Index vector. We call the first component, J, the "index vector". This is defined to be a length-d row vector from $([n] \cup \{\perp\})^d$ in which all non-$\perp$ indices are distinct. On first reading, one should ignore the possibility of $\perp$'s and simply think of an index vector J as an ordered list of d distinct indices from $[n]$.

Diagonalization matrix. We call the second component, A, a "diagonalization matrix". A is $n \times d$ matrix with entries from $\{0, 1, \perp\}$. Most entries in A will be 0; indeed, the row A^j will be nonzero only if j is one of the indices in J. Before describing A completely formally, let us describe the "typical" case when J contains no $\perp$'s, and thus just consists of distinct indices from $[n]$. In this case, A will also contain no $\perp$'s. To make the picture even clearer, let us imagine that J is simply $(1, 2, \ldots, d)$. Thus A will only be nonzero in its first d rows, so let us write $A' = A^{1..d}$. In this case, if $x \in \mathcal{S}$ is the input to `Witness`, then A' will be of the following form:

$$\begin{bmatrix}
\overline{x^1} & * & * & * & \cdots & * \\
x^2 & \overline{x^2} & * & * & \cdots & * \\
x^3 & x^3 & \overline{x^3} & * & \cdots & * \\
x^4 & x^4 & x^4 & \overline{x^4} & \cdots & * \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
x^d & x^d & x^d & x^d & \cdots & \overline{x^d}
\end{bmatrix}.$$

Here each x^j is of course in $\{0, 1\}$, we write $\overline{x^j}$ for $1 - x^j$, and $*$ denotes that the entry may be either 0 or 1. We say that A *diagonalizes x on $J = (1, 2, \ldots, d)$*. We now give the formal definition which includes the possibility of J containing $\perp$'s.

Definition 4.1. Given an index vector J and a solution $x \in \{0, 1\}^n$, we say that the matrix $A \in \{0, 1, \perp\}^{n \times d}$ is a *diagonalization matrix*, and in particular that it *diagonalizes x on J*, if the following conditions hold: If $j \in [n]$ does not appear in J, then row A^j is all zeros. Otherwise, if $j = J_u \in [n]$ for some $u \in [d]$:

$$A^j_t = \perp \text{ if and only } J_t = \perp, \qquad A^j_u = \overline{x^j}, \qquad A^j_t = x^j \text{ for all } t < u \text{ with } J_t \neq \perp. \tag{6}$$

Box list. The last component of a testimony, $\mathcal{B}$, is a list $\mathcal{B} = (B_1, \ldots, B_d)$. For $t \in [d]$ we have that $B_t = \perp$ if $J_t = \perp$, and otherwise B_t is a t-box, as defined in Section 2.3. We define the *dimension* of the box list $\mathcal{B}$ to be $\sum\{t \in [d] : B_t \neq \perp\}$. On first reading, one should ignore the possibility of $B_t = \perp$, in which case $\dim(\mathcal{B})$ is always $1 + 2 + \cdots + d = d(d+1)/2$.

Masking matrix. Having defined the components $(J, A, \mathcal{B})$ of a testimony, we now explain one more piece of notation; that of a masking matrix. Given an index vector J, we define the associated masking matrix $M(J) \in \{0,1\}^{d \times n}$ as follows:

$$M(J)^i_j = \begin{cases} 1 & \text{if } j = J_t \in [n] \text{ for some } t \in [d] \text{ and } i \leq t, \\ 0 & \text{otherwise.} \end{cases}$$

For illustration, if $J = (1, 2, \ldots, d)$, then $M(J)$ is the mostly-zeros $d \times n$ matrix whose left-most $d \times d$ submatrix is

$$\begin{bmatrix} 1 & 1 & 1 & 1 & \ldots & 1 \\ 0 & 1 & 1 & 1 & \ldots & 1 \\ 0 & 0 & 1 & 1 & \ldots & 1 \\ 0 & 0 & 0 & 1 & \ldots & 1 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 1 \end{bmatrix}.$$

Note that in the "typical" case that J contains no $\perp$'s, the number of 1's in $M(J)$ is exactly $d(d+1)/2$. Given a masking matrix, we write $\overline{M(J)}$ for its bitwise complement; i.e., $\overline{M(J)}^i_j = 1 - M(J)^i_j$. We are now able to explain the notation used in the key step of the proof of the Main Theorem. Given the semi-random matrix of values $\boldsymbol{V}$, note that for any J,

$$\boldsymbol{V} = \overline{M(J)} \circ \boldsymbol{V} + M(J) \circ \boldsymbol{V}.$$

Further, the random matrices $\overline{M(J)} \circ \boldsymbol{V}$ and $M(J) \circ \boldsymbol{V}$ are independent of one another. This gives our crucial means of separating the random draw of $\boldsymbol{V}$ into two stages.

4.2 The `Witness` mapping

Here is the deterministic algorithm computing the `Witness` mapping:

`Witness`(x, V):

1. Set $\mathcal{R}_{d+1} = \mathcal{S}$.
2. Initialize J to the length-d column vector $(\perp, \perp, \ldots, \perp)$.
3. Initialize Y to the $n \times d$ matrix where every entry is $\perp$.
4. For $t = d, d-1, d-2, \ldots, 1$:
5. Let $\mathcal{C}_t = \{z \in \mathcal{R}_{t+1} : V^{1..t} z > V^{1..t} x\}$.
6. If $\mathcal{C}_t \neq \emptyset$,
7. Set column Y_t to be the $y \in \mathcal{C}_t$ for which $V^{t+1} y$ is maximal.[†]
8. Set J_t to be the least index in $[n]$ such that $Y_t^{J_t} \neq x^{J_t}$.[‡]
9. Set $\mathcal{R}_t = \{z \in \mathcal{R}_{t+1} : V^{t+1} z > V^{t+1} Y_t \text{ and } z^{J_t} = x^{J_t}\}$.
10. Else
11. Set $\mathcal{R}_t = \mathcal{R}_{t+1}$.
12. End If
13. End For
14. Define the $n \times d$ matrix A by $A_u^j = \begin{cases} Y_u^j & \text{if } j \text{ appears in } J, \\ 0 & \text{otherwise.} \end{cases}$
15. Define the Box list $\mathcal{B} = (B_1, \ldots, B_d)$ as follows:
    For $u \in [d]$, if $J_u = \perp$ then set $B_u = \perp$.
    Otherwise, set B_u to be the u-box containing $Vx - (M(J) \circ V) A_u$.
16. Output $(J, A, \mathcal{B})$.

[†] Two comments about this line: Regarding maximality, say that we break ties by lexicographic order. Regarding the case $t = d$, recall our abuse of notation: $V^{d+1} y$ is defined to be $\mathrm{Obj}^{d+1}(y)$.

[‡] Such an index must exist: $Y^t \neq x$ because $Y^t \in \mathcal{C}_t$ and therefore $V^{1..t} Y^t > V^{1..t} x$.

It is clear that the index vector J and the Box list $\mathcal{B}$ output by `Witness` have the form we claimed. We now verify that `Witness`(x, V) indeed outputs a proper diagonalization matrix A:

Proposition 4.2. *The matrix A output by* `Witness`(x, V) *always diagonalizes x on J.*

Proof: At the end of the algorithm, by definition row A^j is all zeros if j does not appear in J. Thus it remains to analyze each row A^{J_u}, where $u \in [d]$ is such that $J_u \neq \perp$. By definition, we have $A_t^{J_u} = Y_t^{J_u}$ for each $t \in [d]$. Thus we need to verify the three conditions in (6) for $Y_t^{J_u}$. First, $Y_t^{J_u} = \perp$ if and only if Y_t was "not defined" during iteration t of the algorithm (i.e., if $\mathcal{C}_t = \emptyset$), which occurs precisely when $J_t = \perp$. Next, $Y_u^{J_u} = \overline{x^j}$ by definition of J_u. Finally, because of line (9) in `Witness` we have that $z^{J_u} = x^{J_u}$ for $z \in \mathcal{R}_u$. Thus for any $t < u$ where $J_t \neq \perp$ (and thus $\mathcal{C}_t \neq \emptyset$), we have $Y_t^{J_u} = x^{J_u}$ because $Y^{J_u} \in \mathcal{C}_t \subseteq \mathcal{R}_{t+1} \subseteq \mathcal{R}_u$. ■

We also record another simple observation:

Proposition 4.3. *Given an execution of* `Witness`(x, V)*, any two solutions in $\mathcal{R}_t$ have the same J_t'th coordinate, the same J_{t+1}'th coordinate, ..., and the same J_d'th coordinate (excluding the cases $t \leq u \leq d$ where $J_u = \perp$).*

Proof: For a fixed t with $J_t \neq \perp$, the fact that all solutions in $\mathcal{R}_t$ have the same J_t'th coordinate follows immediately from the definition of $\mathcal{R}_t$. The claim for coordinates $J_{t+1}, \ldots, J_d$ follows from the fact that $\mathcal{R}_t \subseteq \mathcal{R}_{t+1} \subseteq \cdots \subseteq \mathcal{R}_d$. ■

This proposition combines with our definition of masking matrices in a crucial way:

Masking Lemma. *Given an execution of* $\texttt{Witness}(x, V)$*, for any* $t \in [d]$ *and* $z, z' \in \mathcal{R}_t$*,*

$$V^t z > V^t z' \quad \Leftrightarrow \quad (\overline{M(J)} \circ V)^t z > (\overline{M(J)} \circ V)^t z'.$$

Proof: We have

$$V^t(z - z') = (\overline{M(J)} \circ V)^t (z - z') + (M(J) \circ V)^t (z - z').$$

By definition of $M(J)$, the row vector $(M(J) \circ V)^t$ has nonzero entries only in indices J_t, $J_{t+1}, \ldots, J_d$ (excluding those J_u's which are $\perp$). But by Proposition 4.3, z and z' agree on these indices. Hence $(M(J) \circ V)^t (z - z') = 0$, and therefore $V^t(z - z') = (\overline{M(J)} \circ V)^t (z - z')$. The lemma follows. ■

Finally, our proof of the key Uniqueness Lemma in Section 5 will rely on the following simpler uniqueness claim:

Testimony–Determines–PO Lemma. *Suppose that we run* $\texttt{Witness}(x, V)$*, where* V *is an outcome for the values such that* x *is Pareto optimal and such that* **OK** *occurs. Then at the end of the run,* x *is uniquely defined by being the* $z \in \mathcal{R}_1$ *with maximal* $V^1 z$*.*

We remark that the assumption that $\mathbf{OK}(V)$ occurs is stronger than necessary; we only need that $V^i y \neq V^i y'$ for all $i \in [d]$ and distinct $y, y' \in \mathcal{S}$ (an event that occurs with probability 1).
Proof: We make the following two claims about the execution of $\texttt{Witness}(x, V)$:

Claim 1: For each $t \in [d+1]$ it holds that x is not t-dominated by any $z \in \mathcal{R}_t$.

Claim 2: x must be in $\mathcal{R}_1$.

Assuming these claims, the lemma follows immediately: $x \in \mathcal{R}_1$ by claim 2, and no $z \in \mathcal{R}_1$ has $V^1 z \geq V^1 x$ by claim 1.

We begin by proving Claim 1. For $t = d + 1$, this follows immediately from the definition of x being Pareto optimal. For smaller t, let us consider the t'th iteration of "For" loop, in which $\mathcal{R}_t$ is defined. We need to consider two cases corresponding to the "If" condition. First suppose $\mathcal{C}_t \neq \emptyset$, so lines (7)—(9) are executed. Now if there were some z in the newly defined $\mathcal{R}_t$ which t-dominated x, then it would satisfy $V^{t+1} z > V^{t+1} Y^t$ and $V^{1..t} z \geq V^{1..t} x$. Since the **OK** event holds, the latter inequality can be strengthened to $V^{1..t} z > V^{1..t} x$. But this means z must be in the set $\mathcal{C}_t$. Since $V^{t+1} z > V^{t+1} Y^t$, we have a contradiction with how Y^t was chosen in line (7). We now consider the second case, that $\mathcal{C}_t = \emptyset$. In this case, $\mathcal{R}_t = \mathcal{R}_{t+1}$. Now by definition of $\mathcal{C}_t = \emptyset$, there is no $z \in \mathcal{R}_{t+1} = \mathcal{R}_t$ which has $V^{1..t} z > V^{1..t} x$. Since the **OK** event occurs, we can strengthen this statement to say that no $z \in \mathcal{R}_t$ can even have $V^{1..t} z \geq V^{1..t} x$, as needed.

We now prove Claim 2. Specifically, we show that $x \in \mathcal{R}_t$ for all $t \in [d+1]$ by (downward) induction on t. The base case, that $x \in \mathcal{R}_{t+1}$, hold by definition. Assume then that $x \in \mathcal{R}_{t+1}$ for some $t \in [d]$. Consider now the t'th iteration of the "For" loop. If the "If" condition does not hold then $\mathcal{R}_t = \mathcal{R}_{t+1} \ni x$, as needed. Assume then that lines (7)—(9) are executed. To show $x \in \mathcal{R}_t$ it suffices to show that $V^{t+1} x > V^{t+1} Y_t$. If this is not true, then $V^{t+1} Y_t \geq V^{t+1} x$, and $V^{1..t} Y_t \geq V^{1..t} x$ also, since $Y_t \in \mathcal{C}_t$. But that means that $Y_t \in \mathcal{R}_{t+1}$ $(t+1)$-dominates x, contradicting Claim 1. ■

5 The Uniqueness Lemma

Let us restate the Uniqueness Lemma.

Uniqueness Lemma. *Fix a testimony $(J, A, \mathcal{B})$ and fix the outcome of $\overline{M(J)} \circ \boldsymbol{V}$. Then there exists at most one $x \in \mathcal{S}$ such that the event*

$$\big(x \in \mathbf{PO}\big) \wedge \big(\texttt{Witness}(x, \boldsymbol{V}) = (J, A, \mathcal{B})\big) \wedge \mathbf{OK} \tag{7}$$

can occur. Here the event's randomness is just the draw of $M(J) \circ \boldsymbol{V}$.

We prove the Uniqueness Lemma in a roundabout way. Specifically, we introduce a second deterministic algorithm called Reconstruct, which takes as input a testimony $(J, A, \mathcal{B})$ and fixed values $\overline{M(J)} \circ V$, and outputs a solution $\underline{x} \in \mathcal{S}$ (or possibly 'FAIL'). Instead of the Uniqueness Lemma as stated, we prove the following:

Uniqueness Lemma′. *Let solution $x \in \mathcal{S}$ and value matrix V be such that x is Pareto optimal and such that event* **OK** *occurs. Assume further that* $\texttt{Witness}(x, V) = (J, A, \mathcal{B})$. *Then* $\underline{\texttt{Reconstruct}}((J, A, \mathcal{B}), (\overline{M(J)} \circ V))$ *outputs x.*

This immediately implies the Uniqueness Lemma, as follows: Fix a testimony $(J, A, \mathcal{B})$ and an outcome $\overline{M(J)} \circ \boldsymbol{V} = \overline{M(J)} \circ V$. Suppose there exist solutions $x, x' \in \mathcal{S}$ for which event (7) can occur (with possibly different outcomes for $M(J) \circ \boldsymbol{V}$). Then Uniqueness Lemma′ tells us that the output of $\underline{\texttt{Reconstruct}}((J, A, \mathcal{B}), (\overline{M(J)} \circ V))$ is both x and x'; hence $x = x'$.

The remainder of this section is devoted to the proof of Uniqueness Lemma′. We begin by defining the algorithm Reconstruct.

$\underline{\texttt{Reconstruct}}((J, A, \mathcal{B}), (\overline{M(J)} \circ V))$:

1. Set $\underline{\mathcal{R}_{d+1}} = \mathcal{S}$.
2. Initialize $\underline{Y}$ to the $n \times d$ matrix where every entry is $\perp$.
3. For $t = d, d-1, d-2, \ldots, 1$:
4. If $J_t \neq \perp$,
5. Write $b \in (\epsilon\mathbb{Z})^t$ for the base point of B_t.
6. Set $\underline{\mathcal{C}'_t} = \{z \in \underline{\mathcal{R}_{t+1}} : (\overline{M(J)} \circ V)^{1..t} z > b$ and $z^j = A_t^j \ \forall$ indices $j \in J\}$.
7. Set $\underline{Y_t}$ to be the $y \in \underline{\mathcal{C}'_t}$ for which $(\overline{M(J)} \circ V)^{t+1} y$ is maximal.*
8. Set $\underline{\mathcal{R}_t} = \{z \in \underline{\mathcal{R}_{t+1}} : (\overline{M(J)} \circ V)^{t+1} z > (\overline{M(J)} \circ V)^{t+1} \underline{Y_t}$ and $z^{J_t} \neq \underline{Y_t}^{J_t}\}$.
9. Else
10. Set $\underline{\mathcal{R}_t} = \underline{\mathcal{R}_{t+1}}$.
11. End If
12. End For
13. Output the $\underline{x} \in \underline{\mathcal{R}_1}$ for which $(\overline{M(J)} \circ V)^1 \underline{x}$ is maximal.

* Some comments about this line. First, if $u = d$ then we interpret $(\overline{M(J)} \circ V)^{d+1} y$ to mean $\text{Obj}^{d+1}(y)$. Second, regarding maximality, we break ties by lexicographic order. Third, for some inputs to Reconstruct it is possible that the set $\underline{\mathcal{C}'_t}$ is empty; in this case one can think of Reconstruct as halting and outputting 'FAIL'. However we will only

be analyzing Reconstruct on inputs where this provably never happen. Finally, the first remark here also applies to line (8) and the second and third remarks here also apply to line (13).

To prove Uniqueness Lemma′, we fix x and V such that x is Pareto optimal and such that event **OK** occurs. We further suppose we have executed $\texttt{Witness}(x, V)$ producing $(J, A, \mathcal{B})$, and then executed $\underline{\texttt{Reconstruct}}((J, A, \mathcal{B}), (\overline{M(J)} \circ V))$ producing $\underline{x}$. Our goal is to show that $\underline{x} = x$. To do this, we will analyze the internal variable assignments made in the executions of Witness and Reconstruct. More specifically, the main task will be to show the following claim asserting that $\underline{\mathcal{R}_t} = \mathcal{R}_t$ for all $t \in [d+1]$. Once we show this, it will be easy to conclude that $\underline{x} = x$ also.

Claim 5.1. *$\underline{\mathcal{R}_t} = \mathcal{R}_t$ for all $t \in [d+1]$.*

Proof: The proof is by (downward) induction. The base case is clear, as $\underline{\mathcal{R}_{d+1}} = \mathcal{R}_d = \mathcal{S}$. For the induction we assume that $\underline{\mathcal{R}_{u+1}} = \mathcal{R}_{u+1}$ for some $u \in [d]$. We now show that $\underline{\mathcal{R}_u} = \mathcal{R}_u$. In doing so, we will also show that $\underline{Y_u} = Y_u$. The set $\underline{\mathcal{C}'_u}$ will not necessarily equal $\underline{\mathcal{C}_u}$, but will be a subset of it.

We henceforth restrict attention to the the $t = u$ iteration of the "For" loop in the execution of Witness and Reconstruct, since this is when variables $\mathcal{R}_u$ and $\underline{\mathcal{R}_u}$ were set. We consider two cases depending on whether or not $J_u = \perp$. In the easy case that $J_u = \perp$, Witness must have assigned $\mathcal{R}_u = \mathcal{R}_{u+1}$, and certainly Reconstruct assigned $\underline{\mathcal{R}_u} = \underline{\mathcal{R}_{u+1}}$. By induction, $\underline{\mathcal{R}_{u+1}} = \mathcal{R}_{u+1}$, and hence $\underline{\mathcal{R}_u} = \mathcal{R}_u$ as required.

The remainder of the claim's proof is devoted to the case that $J_u \neq \perp$, in which case Witness executed its lines (7)—(9) and Reconstruct executed its lines (5)—(8). The B_u referred to in Reconstruct's line (5) is defined at the end of Witness to be u-box containing $Vx - (M(J) \circ V)A_u$. By definition, this means the base point $b \in (\epsilon\mathbb{Z})^u$ used by Reconstruct is such that

$$\begin{aligned} V^{1..u}x - (M(J) \circ V)^{1..u}A_u \;&\in\; b + [0, \epsilon)^u \\ \Rightarrow \qquad V^{1..u}x \;&\in\; \widehat{b} + [0, \epsilon)^u, \\ &\quad \text{where } \widehat{b} = (M(J) \circ V)^{1..u}A_u + b. \end{aligned}$$

Recall that Witness defines $\mathcal{C}_u = \{z \in \mathcal{R}_{u+1} \colon V^{1..u}z > V^{1..u}x\}$. In fact, because we have assumed V causes event **OK** to occur, we may deduce

$$\mathcal{C}_u = \{z \in \mathcal{R}_{u+1} \colon V^{1..u}z > \widehat{b}\}. \tag{8}$$

For if there were some $z \in \mathcal{R}_{u+1}$ and $i \in [u]$ with $\widehat{b}^i < V^i z \leq \widehat{b}^i + \epsilon$, we would have $|V^i z - V^i x| \leq \epsilon$, contradicting the occurrence of **OK**. (The reader may note that this deduction is precisely the reason we introduced the event **OK**.)

Next, recall that Witness defines Y_u to be the $y \in \mathcal{C}_u$ for which $V^{u+1}y$ is maximal (and this maximizer is unique since we assume **OK** occurs). Since $Y_u^j = A_u^j$ for all indices j appearing in J, we must also have that Y_u is the maximizer of $V^{u+1}y$ among all y within the following (nonempty) subset of $\mathcal{C}_u$:

$$\mathcal{C}'_u := \{z \in \mathcal{R}_{u+1} \colon V^{1..u}z > \widehat{b} \text{ and } z^j = A_u^j \text{ for all indices } j \in J\}. \tag{9}$$

Observe that

$$V^{1..u}z > \widehat{b} \quad \Leftrightarrow \quad (\overline{M(J)} \circ V)^{1..u}z + (M(J) \circ V)^{1..u}z > (M(J) \circ V)^{1..u}A_u + b.$$

Since all $z \in \mathcal{C}'_u$ agree with A_u in the indices from J, and since $M(J)$ is nonzero only in columns whose indices are in J, we have that

$$(M(J) \circ V)z = (M(J) \circ V)A_u \quad \text{for every } z \in \mathcal{C}'_u. \tag{10}$$

Therefore an equivalent definition to (9) is

$$\mathcal{C}'_u = \{z \in \mathcal{R}_{u+1} \colon (\overline{M(J)} \circ V)^{1..u} z > b \text{ and } z^j = A^j_u \text{ for all indices } j \in J\}.$$

But $\mathcal{R}_{u+1} = \underline{\mathcal{R}_{u+1}}$ by induction, and hence $\mathcal{C}'_u = \underline{\mathcal{C}'_u}$.

The remainder of the proof of the claim now follows fairly easily using the Masking Lemma from Section 4. Recall that Y_u is the maximizer of $V^{u+1}y$ among all $y \in \mathcal{C}'_u$. On the other hand, `Reconstruct` defines $\underline{Y_u}$ to be the $y \in \underline{\mathcal{C}'_u} = \mathcal{C}'_u$ with maximal $(\overline{M(J)} \circ V)^{u+1}y$. We claim that $\underline{Y_u} = Y_u$. If $u = d$ then this is immediate, as both $V^{d+1}y$ and $(\overline{M(J)} \circ V)^{d+1}y$ are interpreted as $\mathrm{Obj}^{d+1}(y)$. If $u < d$, this follows immediately from the Masking Lemma, using the fact that $\mathcal{C}'_u \subseteq \mathcal{R}_{u+1}$.

Finally, we wish to show that $\underline{\mathcal{R}_u} = \mathcal{R}_u$. Recall that

$$\begin{aligned} & \mathcal{R}_u = \{z \in \mathcal{R}_{u+1} \colon V^{u+1}z > V^{u+1}Y_u \text{ and } z^{J_u} = x^{J_u}\}, \\ \text{and} \qquad & \underline{\mathcal{R}_u} = \{z \in \underline{\mathcal{R}_{u+1}} : (\overline{M(J)} \circ V)^{u+1}z > (\overline{M(J)} \circ V)^{u+1}\underline{Y_u} \text{ and } z^{J_u} \neq \underline{Y_u}^{J_u}\} \\ & \qquad = \{z \in \mathcal{R}_{u+1} : (\overline{M(J)} \circ V)^{u+1}z > (\overline{M(J)} \circ V)^{u+1}Y_u \text{ and } z^{J_u} = x^{J_u}\}; \end{aligned}$$

in this last deduction we used $\underline{\mathcal{R}_{u+1}} = \mathcal{R}_{u+1}$ (by induction), $\underline{Y_u} = Y_u$, and $\overline{Y_u^{J_u}} = x^{J_u}$ (which follows from the definition of $\overline{J_u}$ in `Witness`). If $u = d$ then $\underline{\mathcal{R}_u} = \mathcal{R}_u$ again follows from the interpretation $V^{d+1}z = (\overline{M(J)} \circ V)^{d+1}z = \mathrm{Obj}^{d+1}(z)$. If $u < d$ then $\underline{\mathcal{R}_u} = \mathcal{R}_u$ again follows from the Masking Lemma, noting that $z, Y_u \in \mathcal{R}_{u+1}$. This completes the proof of the induction and hence the claim. ■

Having proven Claim 5.1, it is easy to complete the proof of Uniqueness Lemma′, i.e., to show $\underline{x} = x$. Since the values matrix V is assumed to make x Pareto optimal and make **OK** occur, the Testimony–Determines–PO Lemma from Section 4 implies that x is the solution $z \in \mathcal{R}_1$ with maximal $V^1 z$. On the other hand, $\underline{x}$ is defined to be the solution $z \in \underline{\mathcal{R}_1} = \mathcal{R}_1$ with maximal $(\overline{M(J)} \circ V)^1 z$. But these maximizers are equal by the Masking Lemma.

6 The Boundedness Lemma

In this section we restate and prove the Boundedness Lemma.

Boundedness Lemma. *For every fixed $(J, A, \mathcal{B})$, outcome of $\overline{M(J)} \circ \boldsymbol{V}$, and $x \in \mathcal{S}$, it holds that*

$$\Pr_{M(J) \circ \boldsymbol{V}}[\mathtt{Witness}(x, \boldsymbol{V}) = (J, A, \mathcal{B})] \leq \phi^{\dim(\mathcal{B})} \epsilon^{\dim(\mathcal{B})}.$$

Proof: As in the proof of the Uniqueness Lemma we fix the testimony $(J, A, \mathcal{B})$ and the outcome $\overline{M(J)} \circ \boldsymbol{V} = \overline{M(J)} \circ V$. Unlike the proof of that lemma, we also fix $x \in \mathcal{S}$. By Proposition 4.2 we may assume that matrix A diagonalizes x on J; otherwise the probability of $\mathtt{Witness}(M(J) \circ \boldsymbol{V}) = (J, A, \mathcal{B})$ is 0.

Write $\mathcal{B} = (B_1, \ldots, B_d)$, where each B_t is either a t-box or is $\perp$ (if $J_t = \perp$). For each $t \in [d]$ with $J_t \neq \perp$ we define the event

$$\mathbf{IN}_t = \text{“}\boldsymbol{V}x - (M(J) \circ \boldsymbol{V})A_t \in B_t\text{”},$$

where again, the randomness of these events is just the draw of $M(J) \circ \boldsymbol{V}$. We may complete the proof by showing

$$\Pr_{M(J) \circ \boldsymbol{V}} \left[\bigwedge_{t \in [d] : J_t \neq \perp} \mathbf{IN}_t \right] \leq \phi^{\dim(\mathcal{B})} \epsilon^{\dim(\mathcal{B})}. \tag{11}$$

Recall that

$$M(J)^i_j = \begin{cases} 1 & \text{if } j = J_t \in [n] \text{ for some } t \in [d] \text{ and } i \le t, \\ 0 & \text{otherwise.} \end{cases}$$

We will imagine drawing the random entries of $M(J) \circ \boldsymbol{V}$ in d stages. In the t'th stage we draw the t entries $M(J) \circ \boldsymbol{V}^{1..t}_{J_t}$, unless $J_t = \perp$ in which case we "skip" the t'th stage. By the independence of the entries, the following claim immediately implies (11):

Claim: Assume $t \in [d]$ has $J_t \neq \perp$. Suppose we have completed the first $t-1$ stages of drawing $M(J) \circ \boldsymbol{V}$. Then whether the event $\mathbf{IN}_t$ occurs is determined in the t'th stage, and its probability is at most $\phi^t \epsilon^t$.

To prove the claim we we write $b \in \mathbb{R}^t$ for the base point of B_t and observe that

$$\begin{aligned} \mathbf{IN}_t \quad &\Leftrightarrow \quad \boldsymbol{V}x - (M(J) \circ \boldsymbol{V})A_t \;\in\; B_t \\ \Leftrightarrow \quad (\overline{M(J)} \circ V)^{1..t}x + (M(J) \circ \boldsymbol{V})^{1..t}x - (M(J) \circ \boldsymbol{V})^{1..t}A_t \;&\in\; b + [0,\epsilon)^t \\ \Leftrightarrow \quad (M(J) \circ \boldsymbol{V})^{1..t}(x - A_t) \;&\in\; \left(b - (\overline{M(J)} \circ V)^{1..t}x\right) + [0,\epsilon)^t. \end{aligned} \tag{12}$$

Recalling the definition of $M(J)$ we see that for a fixed $i \in [t]$,

$$(M(J) \circ \boldsymbol{V})^i(x - A_t) = \sum_{i \le u < t : J_u \neq \perp} (M(J) \circ V)^i_{J_u}(x - A_t)^{J_u} \tag{13}$$

$$+ \; (M(J) \circ \boldsymbol{V})^i_{J_t}(x - A_t)^{J_t} \tag{14}$$

$$+ \sum_{u > t : J_u \neq \perp} (M(J) \circ \boldsymbol{V})^i_{J_u}(x - A_t)^{J_u}. \tag{15}$$

Please note that in (13) we have written $M(J) \circ V$ rather than $M(J) \circ \boldsymbol{V}$ because the entries $(M(J) \circ \boldsymbol{V})^i_{J_u}$ for $u < t$ have been fixed prior to the t'th stage. The entries of $M(J) \circ \boldsymbol{V}$ appearing in (14) and (15), however, are still to be drawn.

At this point it may seem as though the event $\mathbf{IN}_t$ as given in (12) depends not only on the entries $(M(J) \circ \boldsymbol{V})^{1..t}_{J_t}$ as stated in the claim, but also on the entries $(M(J) \circ \boldsymbol{V})^{1..t}_{J_u}$ for $u > t$. But this is where we make a crucial observation; indeed, the one which explains why we defined `Witness` to produce diagonalization matrices. By definition of A diagonalizing x on J,

$$(x - A_t)^j = \begin{cases} \pm 1 & \text{if } j = J_t, \\ 0 & \text{if } j = J_u \in [n] \text{ for some } u > t. \end{cases}$$

(If $j = J_u \in [n]$ for some $u < t$ then we cannot say anything about $(x - A_t)^j$, but we do not need to.) Substituting this into (14) and (15), we deduce that

$$(M(J) \circ \boldsymbol{V})^i(x - A_t) = \text{constant} \;\pm (M(J) \circ \boldsymbol{V})^i_{J_t}. \tag{16}$$

In particular, the term (15) has dropped out; hence event (12) does not in fact depend on the entries $(M(J) \circ \boldsymbol{V})^{1..t}_{J_u}$ for $u > t$, as claimed. Finally, substituting (16) into (12) we see that the event $\mathbf{IN}_t$ is equivalent to a conjunction of t events of the form

$$\pm (M(J) \circ \boldsymbol{V})^i_{J_t} \in [c_i, c_i + \epsilon)$$

where the c_i's are fixed constants. Since the random variables $(M(J) \circ \boldsymbol{V})^i_{J_t}$ are independent and have pdf's bounded by ϕ, we conclude that the probability of $\mathbf{IN}_t$ is indeed at most $(\phi\epsilon)^t$, as claimed. ■

7 The Counting Lemma

Here we restate and prove the Counting Lemma.

Counting Lemma. *For a fixed n and ϵ, the quantity*

$$\sum_{\substack{\text{possible testimonies} \\ (J,A,\mathcal{B})}} \phi^{\dim(\mathcal{B})}\epsilon^{\dim(\mathcal{B})} \tag{17}$$

is at most $2 \cdot (4d\phi)^{d(d+1)/2} \cdot n^{2d}$.

Proof: For a given index vector J let us define the following quantities:

$$\text{count}(J) = \#\{t : J_t \neq \perp\}, \quad \text{sum}(J) = \sum\{t : J_t \neq \perp\}, \quad \max(J) = \max\{t : J_t \neq \perp\}.$$

Observe that for a possible testimony $(J, A, \mathcal{B})$, the quantity $\text{sum}(J)$ is identical to $\dim(\mathcal{B})$. We may therefore express (17) as

$$\sum_{\text{possible } J} \phi^{\text{sum}(J)}\epsilon^{\text{sum}(J)} \cdot \#\{(A, \mathcal{B}) \text{ s.t. } (J, A, \mathcal{B}) \text{ is a possible testimony}\}. \tag{18}$$

Let us now count the pairs $(A, \mathcal{B})$ that form possible testimonies with J. By Proposition 4.2 we know that A must diagonalize some solution x on J. There are $2^{\text{count}(J)}$ choices for the values of x^j, for j appearing in J. These force some entries of A; the remaining $\sum\{t-1 : J_t \neq \perp\} = \text{sum}(J) - \text{count}(J)$ entries are free. Thus there are

$$2^{\text{count}(J)}2^{\text{sum}(J)-\text{count}(J)} = 2^{\text{sum}(J)} \text{ possible choices for } A. \tag{19}$$

As for $\mathcal{B}$, let us first count the number of possibilities for $B_{\max(J)}$ (assuming $\max(J)$ exists). We write $m = \max(J)$ for brevity; on first reading, one should think of m as always being d. An execution of $\texttt{Witness}(x, V)$ which is consistent with J and A defines B_m to be the m-box containing the point $p = Vx - (M(J) \circ V)A_m$. Since the entries of V are bounded in $[-1, 1]$ always and since $M(J)$ contains at most d nonzero entries, the point p must lie in $[-n - d, n + d)^d$.[1] There are therefore at most $(2(n + d)/\epsilon)^m$ choices for the box B_m.

We could similarly upper-bound the number of choices for each remaining t-box by $(2(n+d)/\epsilon)^t$; however, this would lead to a final count whose dependence on d was $n^{d+d(d+1)/2}$, rather than n^{2d}. To get the much better dependence of n^{2d} we observe that once B_m is chosen, the remaining t-boxes cannot be "too far away" because, like B_m, they contain a point close to Vx. More precisely, let $t < m$ be such that $J_t \neq \perp$ and consider B_t. It is the t-box containing $\hat{p} = Vx - (M(J) \circ V)A_t$. Now $p - \hat{p} = (M(J) \circ V)(A_t - A_u)$, which means that $\|p - \hat{p}\|_\infty \leq d$. It follows that given the choice of B_m, there are at most $((2d + 1)/\epsilon)^t$ choices for B_t. We conclude that the number of possible choices for $\mathcal{B}$ is at most

$$(2(n + d)/\epsilon)^{\max(J)} \cdot \prod_{t<\max(J):J_t\neq\perp} ((2d + 1)/\epsilon)^t = \left(\frac{2(n + d)}{2d + 1}\right)^{\max(J)} \cdot \left(\frac{2d + 1}{\epsilon}\right)^{\text{sum}(J)}$$
$$\leq \left(\frac{2(n + d)}{2d + 1}\right)^{d} \cdot \left(\frac{2d + 1}{\epsilon}\right)^{\text{sum}(J)}.$$

[1] Proving that p cannot have any coordinate exactly equal to $n + d$ is an exercise for the reader.

Combining this with (19) and substituting into (18), we upper-bound (17) by

$$\sum_{\text{possible } J} \big(2(2d+1)\phi\big)^{\text{sum}(J)} \big(2(n+d)/(2d+1)\big)^d.$$

Finally, we simply upper-bound $\text{sum}(J)$ by $d(d+1)/2$ and the number of possible J by $(n+1)^d$. We conclude that (17) is at most

$$(n+1)^d \big(2(2d+1)\phi\big)^{d(d+1)/2} \big(2(n+d)/(2d+1)\big)^d = (4\phi)^{d(d+1)/2}(d+1/2)^{d(d-1)/2}(n+1)^d(n+d)^d.$$

One may check that $(d+1/2)^{(d-1)/2}(n+1)(n+d) \le 2^{1/d} d^{(d+1)/2} n^2$ for any $d \ge 1$ and $n \ge 3$ (which we may assume, as our final bound is always at least 2^3). Hence (17) is indeed at most

$$2(4d\phi)^{d(d+1)/2} n^{2d},$$

as claimed. ■

8 Conclusion

There are several open problems that remain. One intriguing problem is to show a lower bound for the expected number of Pareto optima in which the exponent on n grows with d. Currently we cannot rule out the possibility of an upper bound of the form $f(d,\phi)n^2$; however we regard this possibility as unlikely. We feel it is likely that there is a lower bound of at least $\Omega(n^d)$ for constant d and ϕ; our intuition is partly based on the known lower bound of $\Omega(n^d)$ in the scenario of 2^n completely independent points uniformly distributed on $[-1,1]^{d+1}$.

Another interesting open problem is whether our methods can be used to give improved upper bounds on the higher moments of the number of Pareto optima in the smoothed analysis model. This is currently unclear; we know of no bounds that improve on those of Röglin and Teng [RT09]. Finally, one could ask about reducing the factor of $(\phi d)^{d(d+1)/2}$ in our bound, as well as whether our results extend to the case of solutions in $\{0,1,2,\ldots,c\}^n$ for integer constants $c > 1$.

8.1 Acknowledgements

Part of this research was done during a visit to Microsoft Research New England; we thank them for their hospitality. We also thank Shang-Hua Teng and Ilias Diakonikolas for sharing their expertise.

References

[ANRV07] Heiner Ackermann, Alantha Newman, Heiko Röglin, and Berthold Vöcking. Decision-making based on approximate and smoothed Pareto curves. *Theoretical Computer Science*, 378(3):253–270, 2007. 1.3

[Bei04] René Beier. *Probabilistic Analysis of Discrete Optimization Problems*. PhD thesis, Universität des Saarlandes, 2004. 1.3

[BKS01] Stephan Börzsöny, Donald Kossmann, and Konrad Stocker. The Skyline operator. In *Proceedings of the 17th Annual International Conference on Data Engineering*, pages 421–430, 2001. 1.1

[BKST78] Jon Bentley, Hsiang-Tsung Kung, Mario Schkolnick, and Clark Thompson. Random knapsack in expected polynomial time. *Journal of the ACM*, 25(4):536–543, 1978. 1.4

[BRV07] René Beier, Heiko Röglin, and Berthold Vöcking. The smoothed number of Pareto optimal solutions in bicriteria integer optimization. In *Proceedings of the 11th Annual Conference on Integer Programming and Combinatorial Optimization*, pages 53–67, 2007. 1.3, 2.3

[Buc89] Christian Buchta. On the average number of maxima in a set of vectors. *Information Processing Letters*, 33(2):63–65, 1989. 1.4

[BV04] René Beier and Berthold Vöcking. Random knapsack in expected polynomial time. *Journal of Computer and System Sciences*, 69(3):306–329, 2004. 1.1, 1.2, 1.4

[BV06] René Beier and Berthold Vöcking. Random knapsack in expected polynomial time. *Typical Properties of Winners and Losers in Discrete Optimization*, 35(4):855–881, 2006. 1.3, 1.4

[Deb01] Kalyanmoy Deb. *Multi-objective optimization using evolutionary algorithms*. Wiley, 2001. 1.1

[Dev80] Luc Devroye. A note on finding convex hulls via maximal vectors. *Information Processing Letters*, 11(1):53–56, 1980. 1.4

[DF89] Martin Dyer and Alan Frieze. Probabilistic analysis of the multidimensional knapsack problem. *Mathematics of Operations Research*, 14(1):162–176, 1989. 1.2

[Dia10] Ilias Diakonikolas. *Approximation of Multiobjective Optimization Problems*. PhD thesis, Columbia University, 2010. 1.1

[Ehr05] Matthias Ehrgott. *Multicriteria optimization*. Springer, 2005. 1.1

[GMS84] Andrew Goldberg and Alberto Marchetti-Spaccamela. On finding the exact solution of a zero-one knapsack problem. In *Proceedings of the 16th Annual ACM Symposium on Theory of Computing*, pages 359–368, 1984. 1.2

[Lue98] George Lueker. Average-case analysis of off-line and on-line Knapsack problems. *Journal of Algorithms*, 29(2):277–305, 1998. 1.2

[NU69] George Nemhauser and Zev Ullmann. Discrete dynamic programming and capital allocation. *Management Science*, 15(9):494–505, 1969. 1.1

[PY02] Christos Papadimitriou and Mihalis Yannakakis. On the approximability of trade-offs and optimal access of web sources. In *Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science*, pages 86–92, 2002. 1.1

[RT09] Heiko Röglin and Shang-Hua Teng. Smoothed analysis of multiobjective optimization. In *Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science*, pages 681–690, 2009. 1.3, 1.4, 8

[RV07] Heiko Röglin and Berthold Vöcking. Smoothed analysis of integer programming. *Mathematical Programming*, 110(1):21–56, 2007. 1.3

[ST04] Daniel 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. 1.2

[Ten10] Shang-Hua Teng, 2010. National Science Foundation award #0964481. Abstract available at `http://www.nsf.gov/awardsearch/showAward.do?AwardNumber=0964481`. 1.3