THEORY OF COMPUTING, Volume 10 (10), 2014, pp. 237–256
www.theoryofcomputing.org

Lower Bounds for the Average and Smoothed Number of Pareto-Optima

Tobias Brunsch   Navin Goyal   Luis Rademacher   Heiko Röglin

Received February 4, 2013; Revised May 13, 2014; Published September 29, 2014

Abstract: Smoothed analysis of multiobjective 0–1 linear optimization has drawn considerable attention recently. The goal is to give bounds for the number of Pareto-optimal solutions (i. e., solutions with the property that no other solution is at least as good in all the coordinates and better in at least one) for multiobjective optimization problems. In this article we prove several lower bounds for the expected number of Pareto optima. Our basic result is a lower bound of $\Omega_d(n^{d-1})$ for optimization problems with d objectives and n variables under fairly general conditions on the distributions of the linear objectives. Our proof relates the problem of finding lower bounds for the number of Pareto optima to results in discrete geometry and geometric probability about arrangements of hyperplanes. We use our basic result to derive the following results: (1) To our knowledge, the first lower bound for natural multiobjective optimization problems. We illustrate this on the maximum spanning tree problem with randomly chosen edge weights. Our technique is sufficiently flexible to yield such lower bounds also for other standard objective functions studied in this setting (such as multiobjective shortest path, TSP, matching). (2) A smoothed lower bound of $\min\{\Omega_d(n^{d-1.5}\phi^d), 2^{\Theta_d(n)}\}$ for ϕ-smooth instances of the 0–1 knapsack problem with d profits.

ACM Classification: F.2.2, G.1.6, G.3

AMS Classification: 68Q25

Key words and phrases: multiobjective optimization, probabilistic analysis, smoothed analysis

Preliminary versions of parts of this paper appeared in the Proceedings of the 8th Annual Conference on Theory and Applications of Models of Computation (TAMC'11) [6] and the Proceedings of the 32nd Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS'12) [13].

DOI: 10.4086/toc.2014.v010a010

1 Introduction

Multiobjective optimization involves scenarios where there is more than one objective function to optimize: when planning a train trip we may want to choose connections that minimize fare, total time, number of train changes, etc. The objectives may be in conflict with each other and there may not be a single best solution to the problem. Such multiobjective optimization problems arise in diverse fields ranging from economics to computer science, and have been well-studied. A number of approaches exist in the literature to deal with the trade-offs among the objectives in such situations: goal programming, multiobjective approximation algorithms, Pareto-optimality; see, e. g., [14, 15, 19] for references. It is the latter approach using Pareto-optimality that concerns us in this article. A Pareto-optimal solution is a solution with the property that no other solution is at least as good in all the objectives and better in at least one. Clearly, the set of Pareto-optimal solutions (Pareto set in short) contains all desirable solutions as any other solution is strictly worse than a solution in the Pareto set. In the worst case, the Pareto set can be exponentially large as a function of the input size (see, e. g., [11, 16]). However, in many cases of interest, the Pareto set is typically not too large. If the Pareto set is small and can be generated efficiently, then often some possibly human-assisted post-processing is applied to make a choice among the Pareto-optimal solutions after the Pareto set has been generated. Pareto sets are also used in heuristics for optimization problems (e. g., [17]). To explain why Pareto sets are frequently small in practice, multiobjective optimization has recently been studied from the view-point of smoothed analysis [22]. We introduce some notation before describing this work.

Notation. For a positive integer n, we denote the set $\{1,2,\dots,n\}$ by $[n]$. We will use the partial order $\preceq$ in $\mathbb{R}^d$ defined by $x \preceq y$ iff for all $i \in [d]$ we have $x_i \leq y_i$. For $a,b \in \mathbb{R}^d$ we say that b dominates a if $b_i \geq a_i$ for all $i \in [d]$, and for at least one $i \in [d]$, we have strict inequality. We denote the relation "b dominates a" by $b \succ a$.

We say that a probability distribution over $\mathbb{R}^d$ is *symmetric* if for any measurable set A we have that the probability of A is the same as the probability of $-A$. Recall that a distribution is *absolutely continuous* if it has a density function. We say that a random variable is symmetric (or absolutely continuous) if its distribution is symmetric (or absolutely continuous, respectively).

The multiobjective optimization problems we study have binary variables and linear objective functions. In a general setting, the feasible solution set is an arbitrary set $\mathcal{S} \subseteq \{0,1\}^n$. The problem has d linear objective functions $v^{(i)} : \mathcal{S} \to \mathbb{R}$, given by

$$v^{(i)}(x) = \sum_{j \in [n]} v_j^{(i)} x_j\,, \qquad \text{for } i \in [d],$$

and $(v_1^{(i)},\dots,v_n^{(i)}) \in \mathbb{R}^n$ (so $v^{(i)}$ is also interpreted as an n-dimensional vector in the natural way). For convenience, we will assume, unless otherwise specified, that we want to maximize all the objectives, and we will refer to the objectives as profits. This entails no loss of generality. Thus the optimization problem considered is the following.

$$\begin{aligned} &\text{maximize } v^{(1)}(x),\dots,\text{maximize } v^{(d)}(x), \\ &\text{subject to } x \in \mathcal{S}. \end{aligned} \tag{1.1}$$

Let V be the $d \times n$ matrix with rows $v^{(1)}, \ldots, v^{(d)}$. A solution $x \in \mathcal{S}$ is said to be Pareto-optimal (or maximal under $\preceq$) if $Vx \not\prec Vy$ for all $y \in \mathcal{S}$. For a set X of points in $\mathbb{R}^n$, let $p(X)$ denote the number of Pareto optima in X.

Multiobjective knapsack. For the special case of the multiobjective 0–1 knapsack problem we have $d+1$ linear objective functions. One of the objective functions, $w : \{0,1\}^n \to \mathbb{R}$, called weight, is given by

$$w(x) = \sum_{j \in [n]} w_j x_j\,.$$

Weight is to be minimized. The other d objective functions $v^{(i)} : \{0,1\}^n \to \mathbb{R}$ are profits as before and are given by $v^{(i)}(x) = \sum_{j \in [n]} v_j^{(i)} x_j$. Profits are to be maximized. We require that all the entries in w and $v^{(i)}$ come from $[0,1]$.

For the knapsack problem, we need to modify the definition of domination appropriately because while we want to maximize the profit, we want to minimize the weight. It will be clear from the context which notion is being used.

For the knapsack problem, we need to modify the definition of domination appropriately because while we want to maximize the profit, we want to minimize the weight. It will be clear from the context which notion is being used.

Smoothed analysis. For our multiobjective optimization problem (1.1), in the worst case the size of the Pareto set can be exponential even for $d = 2$ (the bicriteria case). Smoothed analysis is a framework for the analysis of algorithms introduced by Spielman and Teng [22] to explain the fast running time of the Simplex algorithm in practice, despite having exponential running time in the worst case. Since its introduction, smoothed analysis has been applied to a variety of algorithms. Beier and Vöcking [3] studied the 0–1 knapsack problem under smoothed analysis. In our context of multiobjective optimization, smoothed analysis would mean that the instance (specified by V) is chosen adversarially, but then each entry is independently perturbed according to, say, Gaussian noise with small standard deviation. In fact, Beier and Vöcking [3] introduced a more general notion of smoothed analysis. In one version of their model, each entry of the matrix V is an independent random variable taking values in $[-1,1]$ with the restriction that each has probability density function bounded from above by ϕ, for a parameter $\phi \geq 1$. We refer to distributions supported on $[-1,1]$ with probability density bounded above by ϕ as ϕ-smooth distributions. This model is more general than Spielman and Teng's because, by choosing the densities appropriately, the adversary can not only determine the mean values of the entries, as in the original model, but also the type of noise. For greater generality, one of the rows of V could be chosen fully adversarially (deterministically). As ϕ is increased, the smoothed model becomes more like the worst-case model. With the exception of Theorem 1.4 below, we do not require adversarial choice of a row in V.

Previous work. Beier and Vöcking [2] showed that in the above model for the 0–1 knapsack problem with adversarial weights the expected number of Pareto optima is $O(n^4\phi)$. The result generalizes to other bicriteria optimization problems. Beier et al. [1] make this generalization explicit and improve the upper bound to $O(n^2\phi)$. Röglin and Teng [19] studied multiobjective optimization problems in the above framework. They showed that the expected size of the Pareto set with one adversarial and d ϕ-smooth objectives is of the form $O_d((n\phi)^{2^{d-2}(d+1)!})$. Note that with the notation O_d, and, analogously with Ω_d and Θ_d we suppress factors that depend only on d. Moitra and O'Donnell [15] improved this upper bound to $2n^{2d}\cdot(4\phi d)^{d(d+1)/2}$. (Again, these results allow one of the objectives to be chosen adversarially.) Very recently, this has been improved to $O_d(n^{2d}\phi^d)$ for the mildly restricted class of quasiconcave ϕ-smooth instances [7]. The question of a lower bound for the expected number of Pareto optima was raised in [19] and [15].

An early average-case lower bound of $\Omega(n^2)$ was proved in [2] for the knapsack problem with a single profit vector. This result, however, required an adversarial choice of exponentially increasing weights.

Ehrgott [11] presents examples showing that many problems (including the 0–1 knapsack problem and the maximum spanning tree problem) can have an exponential number of Pareto optima in the worst case even for only two objective functions.

Our results. In this article we prove lower bounds for the expected number of Pareto optima. Our basic result, Theorem 1.1, deals with the case when the entries of the matrix V are independent, symmetrically distributed, absolutely continuous random variables. Note that we do not require that the distributions be identical: each entry can have a different distribution. This generality will in fact be useful in our lower bound for the maximum spanning tree problem, Theorem 1.2. Note that all entries of V are random, unlike in the results discussed above where one of the objectives is chosen adversarially. This makes our lower bound stronger.

Theorem 1.1 (Basic theorem)**.** *Suppose all entries of a $d\times n$ random matrix V are independent, symmetrically distributed, absolutely continuous random variables. Let X denote the random set $\{Vr : r\in\{0,1\}^n\}$. Then*

$$\mathop{\mathbb{E}}_{V} p(X) \geq \frac{1}{2^{d-1}}\sum_{k=0}^{d-1}\binom{n-1}{k}. \tag{1.2}$$

This implies the simpler bound

$$\mathop{\mathbb{E}}_{V} p(X) \geq \left(\frac{n-1}{2(d-1)}\right)^{d-1}$$

using the well-known bound

$$\binom{n-1}{d-1} \geq \left(\frac{n-1}{d-1}\right)^{d-1}.$$

We give two proofs of this result. The two proofs have a similar essence, but different form. Both proofs relate the problem at hand to some well-known results in geometry. This connection with geometry is new and may be useful for future research. The first proof lower bounds the expected number of Pareto optima of a point set by the expected number of vertices of its convex hull (up to a constant that depends on d but not on n) and then invokes known lower bounds for the expected number of vertices of projections of hypercubes, as the random set in Theorem 1.1 is a linear image of the vertices of the hypercube. The second proof gives a characterization of maximality in terms of 0–1 vectors and then relaxes integrality to get a relaxed dual characterization by means of convex separation, which reduces counting Pareto optima to giving a lower bound on the probability that the convex hull of n random points contains the origin. This probability is known exactly by a theorem of Wendel (Theorem 2.2).

Interestingly, our lower bound is basically the same as the expected number of Pareto optima when 2^n uniformly random points are chosen from $[-1,1]^d$, which is shown to be $\Theta_d(n^{d-1})$ in several papers [4, 9, 8]. This raises the possibility of a closer connection between the two models; such a connection could be useful as the model of uniformly random points is understood better.

The basic theorem above corresponds to the case when the set of feasible solutions $\mathcal{S}$ is $\{0,1\}^n$. But in many interesting cases $\mathcal{S}$ is a strict subset of $\{0,1\}^n$: for example, in the multiobjective spanning tree problem, n is the number of edges in an underlying network, and $\mathcal{S}$ is the set of incidence vectors of spanning trees in the network; similarly, for the multiobjective shortest path problem, $\mathcal{S}$ is the set of incidence vectors of s–t paths. We can use our basic theorem to prove lower bounds for the size of the Pareto set for such $\mathcal{S}$. Our technique is pliable enough to give interesting lower bounds for many standard objective functions used in multiobjective optimization (in fact, any standard objective that we tried): multiobjective shortest path, TSP, matching, arborescence, etc. We will illustrate the idea with the multiobjective spanning tree problem on the complete graph. In this problem, we have the complete undirected graph K_n on n vertices as the underlying graph. Each edge e has a set of profits $v^{(i)}(e) \in [-1,1]$ for $i \in [d]$. The set $\mathcal{S}$ of feasible solutions is given by the incidence vectors of spanning trees of K_n. Notice that the feasible solutions here live in $\{0,1\}^{\binom{n}{2}}$ and not in $\{0,1\}^n$.

Theorem 1.2. *In the d-objective maximum spanning tree problem on K_n there exists a choice of 2-smooth distributions such that the expected number of Pareto-optimal spanning trees is at least*

$$\left(\frac{n-3}{2(d-1)}\right)^{d-1}.$$

The proof of this theorem utilizes the full power of Theorem 1.1, namely the ability to choose different symmetric distributions.

In our basic theorem above, Theorem 1.1, we required the distributions to be symmetric, and therefore that theorem does not apply to the 0–1 knapsack problem where all profits and weights are non-negative. With a slight loss in the lower bound we also get a lower bound for this case. In the d-dimensional 0–1 knapsack problem we have d objectives $v^{(i)}$ for $i \in [d]$, called profits, and an additional objective w, called weight. Components of $p^{(i)}$ and w are all chosen from $[0,1]$. We want to maximize the profits and minimize the weight, and so the definitions of domination and Pareto-optimality are accordingly modified.

Theorem 1.3. *For the multiobjective 0–1 knapsack problem where all the weight components are 1 and profit components are chosen uniformly at random from $[0,1]$, the expected number of Pareto optima is $\Omega_d(n^{d-1.5})$.*

Theorem 1.1 or Theorem 1.3 (depending on whether one wants a bound for non-negative or unrestricted weights and profits) can be used to give the following lower bound on the expected number of Pareto optima when the profits are ϕ-smooth (actually, uniform in carefully chosen intervals of length at least $1/\phi$):

Theorem 1.4. *For any fixed $d \geq 2$ and for $n \geq 8d$ and $\phi \geq 2d$ there exist*

1. *weights $w_1, \dots, w_n \geq 0$,*
2. *intervals $[a_{ij}, b_{ij}] \subseteq [0,1]$, $i \in [d]$, $j \in [n]$ of length at least $1/\phi$ and with $a_{ij} \geq 0$, and*
3. *a set $\mathcal{S} \subseteq \{0,1\}^n$*

such that if profits $v_j^{(i)}$ are chosen independently and uniformly at random in $[a_{ij}, b_{ij}]$, then the expected number of Pareto-optimal solutions of the d-dimensional knapsack problem with solution set $\mathcal{S}$ is at least

$$\min\left\{\Omega_d(n^{d-1.5}\phi^d), 2^{\Theta_d(n)}\right\}.$$

This lower bound should be contrasted with the aforementioned upper bounds $2 \cdot (4\phi d)^{d(d+1)/2} \cdot n^{2d}$ [15] and $O_d(n^{2d}\phi^d)$ [7]. This latter bound is for the mildly restricted but natural class of ϕ-smooth instances with quasiconcave densities, where a density f is said to be quasiconcave if there exists a value $x \in \mathbb{R}$ such that it is non-decreasing below x and non-increasing above x. The densities in our lower bound instances are quasiconcave. For general multiobjective optimization (basically without the restriction of entries being non-negative) the exponent of n becomes exactly $d-1$ in Theorem 1.4.

Theorem 1.4 requires $\mathcal{S}$ to be chosen adversarially. To our knowledge, no non-trivial lower bounds were known before our work for natural choices of $\mathcal{S}$. This is addressed by our Theorem 1.1 ($\mathcal{S} = \{0,1\}^n$) and Theorem 1.2 ($\mathcal{S}$ is the set of spanning trees of the complete graph) above; these theorems are for a small constant value of ϕ, and therefore do not clarify what the dependence on ϕ should be.

2 The basic theorem

In this section we prove Theorem 1.1. We will include two proofs that, while in essence the same, emphasize the geometric and algebraic views, respectively. Also the second proof is more self-contained. It is perhaps worth mentioning that we first discovered the second proof, and in the course of writing the present article we found the ideas and known results that can be combined to get a more geometric proof.

2.1 First proof

We will use the following result that relates the number of Pareto optima of a point set with the number of vertices of its convex hull:

Theorem 2.1 ([4], [18, Theorem 4.7])**.** *Let P be a finite subset of $\mathbb{R}^d$ and for $\sigma \in \{-1,+1\}^d$ let*

$$P_\sigma = \left\{(\sigma_1 x_1, \dots, \sigma_d x_d) : (x_1, \dots, x_d) \in P\right\}.$$

For every vertex $(x_1, \dots, x_d)$ of the convex hull of P there exists $\sigma \in \{-1,+1\}^d$ such that $(\sigma_1 x_1, \dots, \sigma_d x_d)$ is maximal under $\preceq$ with respect to the set P_σ.

First proof of Theorem 1.1. The convex hull of X is a random polytope. By Theorem 2.1, every vertex is maximal under our partial order $\preceq$ for at least one of the 2^d reflections involving coordinate hyperplanes. This means that

$$|\text{vertices of conv}\, X| \leq \sum_{\sigma\in\{-1,1\}^d} p(X \text{ with coordinates of points flipped by signs in } \sigma)\,, \tag{2.1}$$

where $\operatorname{conv} X$ denotes the convex hull of X. Our symmetry assumption followed by (2.1) implies

$$\begin{aligned}\mathop{\mathbb{E}}_V(p(X)) &= \frac{1}{2^d}\cdot\sum_{\sigma\in\{-1,1\}^d}\mathop{\mathbb{E}}_V p(X \text{ with coordinates of points flipped by signs in } \sigma)\\ &\geq \frac{1}{2^d}\cdot\mathop{\mathbb{E}}_V|\text{vertices of conv}\, X|\,.\end{aligned}$$

This idea is from [4]. It is used there in the opposite direction, that is, to get upper bounds for the expected number of vertices from upper bounds for the expected number of maximal points.

In order to count the vertices of the convex hull of X, we observe that the convex hull of X has a special structure: it is the image of an n-dimensional hypercube via a linear map. Such an image is called a *zonotope*. It is clear that a set in $\mathbb{R}^d$ is a zonotope iff it is the Minkowski sum of a finite number of segments. It is known [10, Theorem 1.8] that for a matrix V with columns in general position (that is, any d columns are linearly independent, which happens almost surely (with probability 1) in our case), the number of vertices of $\operatorname{conv} X$ is equal to the maximum number of vertices of a d-dimensional zonotope formed as the Minkowski sum of n line segments, and this number is known exactly [12, 31.1.1]. That is, almost surely,

$$|\text{vertices of conv}\, X| = 2\sum_{k=0}^{d-1}\binom{n-1}{k}\,.$$

The claimed bound follows. ☐

2.2 Second proof

Some more definitions before getting into the proof: set $\mathbb{R}_+ = \{x\in\mathbb{R} : x\geq 0\}$ and $\mathbb{R}_- = \{x\in\mathbb{R} : x\leq 0\}$. For $\sigma\in\{-1,1\}^d$, the orthant associated with σ is $\{(\sigma_1x_1,\dots,\sigma_dx_d) : (x_1,\dots,x_d)\in\mathbb{R}_+^d\}$. In particular, if σ is the all 1's vector then we call its associated orthant the positive orthant, and if σ is the all -1's vector then we call its orthant the negative orthant. For a finite set of points $P=\{p_1,\dots,p_k\}\subseteq\mathbb{R}^d$, the conic hull is denoted $\mathsf{cone}(P) = \{\sum_{i=1}^k \alpha_i p_i : \alpha_i\geq 0\}$ (note that the conic hull is always convex).

We will use the following result by Wendel.

Theorem 2.2 ([23], [21, Theorem 8.2.1]). *If $X_1,\dots,X_n$ are independent random points in $\mathbb{R}^d$ whose distributions are symmetric (but not necessarily identical) and such that with probability 1 all subsets of size d are linearly independent, then*

$$\Pr[0\notin\operatorname{conv}\{X_1,\dots,X_n\}] = \frac{1}{2^{n-1}}\sum_{k=0}^{d-1}\binom{n-1}{k}\,.$$

The linear independence condition holds in particular under the simpler assumption that no hyperplane through the origin is assigned positive probability by any of the n distributions. For example, it holds when the points are drawn i. i. d. at random from the unit sphere.

Second proof of Theorem 1.1. By linearity of expectation we have $\mathbb{E}\, p(X) = \sum_r \Pr[Vr \text{ maximal}]$. Notice that $\Pr[Vr \text{ maximal}]$ does not depend on r because all entries of V are symmetrically distributed, so we can write $\mathbb{E}\, p(X) = 2^n \Pr[V\mathbf{1} \text{ maximal}]$, where $\mathbf{1}$ denotes the all 1's vector.

For the rest of the proof we will focus on finding a lower bound for this last probability. To understand this probability we first rewrite the event $[V\mathbf{1} \text{ maximal}]$ in terms of a different event via easy intermediate steps:

$$[V1 \text{ maximal}] = [Vr \not> V\mathbf{1},\ \forall r \in \{0,1\}^n] = [0 \not> V(\mathbf{1}-r),\ \forall r \in \{0,1\}^n] = [0 \not> Vr,\ \forall r \in \{0,1\}^n]\,.$$

Now we have

$$\Pr\big[0 \not> Vr,\ \forall r \in \{0,1\}^n\big] \geq \Pr\big[0 \not> Vr,\ \forall r \in [0,1]^n\big]\,.$$

Event $\big[0 \not> Vr,\ \forall r \in [0,1]^n\big]$ is the same as the event $[\mathsf{cone}(v_1,\dots,v_n) \cap \mathbb{R}^d_- = \{0\}]$. That is to say, the cone generated by the non-negative linear combinations of $v_1,\dots,v_n$ does not have a point distinct from the origin that lies in the negative orthant.

Then, by the separability property of convex sets, there exists a hyperplane $H = \{x \in \mathbb{R}^d : \langle u,x\rangle = 0\}$ separating $\mathsf{cone}(v_1,\dots,v_n)$ and $\mathbb{R}^d_-$. More precisely, there exists a vector $u \in \mathbb{R}^d_+ \setminus \{0\}$ such that $\mathsf{cone}(v_1,\dots,v_n)\cdot u \geq 0$ and this implies

$$\Pr[\mathsf{cone}(v_1,\dots,v_n) \cap \mathbb{R}^d_- = \{0\}] = \Pr[\exists u \in \mathbb{R}^d_+ \setminus \{0\} : \mathsf{cone}(v_1,\dots,v_n)\cdot u \geq 0]\,.$$

Now

$$\begin{aligned}
&\Pr[\mathsf{cone}(v_1,\dots,v_n) \text{ contained in some halfspace}] \\
&\qquad \leq \sum_{\sigma\in\{-1,1\}^d} \Pr[\mathsf{cone}(v_1,\dots,v_n) \text{ contained in some halfspace with inner normal in orthant } \sigma] \\
&\qquad = 2^d \Pr[\exists u \in \mathbb{R}^d_+ \setminus \{0\} : \mathsf{cone}(v_1,\dots,v_n)\cdot u \geq 0]\,.
\end{aligned}$$

Clearly, we have

$$[\mathsf{cone}(v_1,\dots,v_n) \text{ contained in some halfspace}] = [v_1,\dots,v_n \text{ contained in some halfspace}]\,.$$

Theorem 2.2 and the fact that the distribution of v_i is symmetric and assigns measure zero to every hyperplane through 0 imply

$$\Pr[v_1,\dots,v_n \text{ contained in some halfspace}] = \Pr[0 \notin \operatorname{conv}\{v_1,\dots,v_n\}] = \frac{1}{2^{n-1}}\sum_{k=0}^{d-1}\binom{n-1}{k}\,.$$

We conclude:

$$\mathbb{E}\, p(X) \geq \frac{1}{2^{d-1}}\sum_{k=0}^{d-1}\binom{n-1}{k}\,. \qquad \square$$

The following easy corollaries of Theorem 1.1 will be useful later.

Corollary 2.3. *Theorem 1.1 holds also for the feasible set* $\mathcal{S} = \{-1,1\}^n$.

Proof. For $x \in \{0,1\}^n$ define $x' = (2x_1 - 1, 2x_2 - 1, \ldots, 2x_n - 1) = 2x - \mathbf{1}$, where $\mathbf{1}$ is the all 1's vector. Thus $x' \in \{-1,1\}^n$. Now $Vx' = 2Vx - V\mathbf{1}$ and so $Vx' \preceq Vy'$ if and only if $Vx \preceq Vy$. Hence x is Pareto-optimal for the feasible set $\{0,1\}$ if and only if x' is Pareto-optimal for the feasible set $\{-1,1\}$. This immediately implies the corollary. $\square$

Corollary 2.4. *Under the assumptions of Theorem 1.1 and when the set of feasible solutions is* $\mathcal{S} \subseteq \{0,1\}^n$ *we have*

$$\mathbb{E}\, p(X) \geq \frac{|\mathcal{S}|}{2^{n+d-1}} \sum_{k=0}^{d-1} \binom{n-1}{k}. \tag{2.2}$$

Proof. For any given $r \in \mathcal{S}$, the probability that it is Pareto-optimal in the current instance with solution set restricted to $\mathcal{S}$ is at least the probability that it is Pareto-optimal in the instance with solution set $\{0,1\}^n$. By the symmetry of the distributions this probability is independent of r and by Theorem 1.1 it is at least

$$\frac{1}{2^{n+d-1}} \sum_{k=0}^{d-1} \binom{n-1}{k}.$$

Linearity of expectation completes the proof. $\square$

3 Lower bound for multiobjective maximum spanning trees

In this section we show that our basic result can be used to derive similar lower bounds for $\mathcal{S}$ other than those encountered earlier in this article. We illustrate this for the case of the multiobjective maximum spanning tree problem on the complete graph; for this problem, $\mathcal{S}$ is the set of incidence vectors of all spanning trees on n vertices. The idea of the proof is simple: "embed" the given instance of the basic problem into an instance of the problem at hand. The proof requires the full power of Theorem 1.1, namely the distributions of the entries of V need not be identical. It is worth noting that the direct use of Corollary 2.4 does not provide any useful bound for the case of spanning trees because in this case the right hand side of (2.2) becomes less than 1. The proof below can easily be modified so that all profits are chosen from intervals of non-negative numbers.

We now prove Theorem 1.2.

Proof of Theorem 1.2. The idea of the proof is to embed an instance of the case $\mathcal{S} = \{0,1\}^{n-2}$ of the basic theorem into the tree instance. We now describe our tree instance. We identify a subgraph G of K_n (the complete graph on n vertices): the vertex set of G is the same as the vertex set of K_n, which we denote by $\{s,t,u_1,u_2,\ldots,u_{n-2}\}$ for convenience. The edge set of G consists of the edge (s,t), and edges $(s,u_j),(t,u_j)$. Thus, G consists of $2n-3$ edges. Now we choose the distribution of the profits for each edge of K_n. For edges outside G, the distribution for all profits is identical and it is simply the uniform distribution on $[-1,-1/2]$. For edge (s,t), the distribution is uniform over $[1/2,1]$. For all other edges (i. e., the edges (s,u_j) and (t,u_j)), it is uniform over $[-1/2,1/2]$. Note that these densities are not symmetric. This is not a problem because we will see below that we do not need to apply Theorem 1.1 directly to these densities, but only to symmetric densities that are derived from them. Let $\mathcal{T}$ denote the set of spanning trees that include edge (s,t), and for every other vertex u_j, exactly one of (s,u_j) and (t,u_j). Clearly $|\mathcal{T}| = 2^{n-2}$. The result of the above choices of distributions is that all the Pareto-optimal spanning trees come from $\mathcal{T}$:

Claim 3.1. *For any choices of profits from the intervals as specified above, if a tree T is Pareto-optimal then $T \in \mathcal{T}$.*

Proof of the claim. Fix any choice of profits as above. Suppose that a tree T' is Pareto-optimal but $T' \notin \mathcal{T}$. Then: either (1) T' has an edge e outside $E(G)$, or (2) all its edges are from $E(G)$ but it does not use edge (s,t). In case (1), remove the edges from T' that are not in $E(G)$ (these have profits at most $-1/2$), and then complete the remaining disconnected graph to a spanning tree using edges from $E(G)$ (these have profit at least $1/2$). Clearly, the resulting tree is heavier than T' in each of the d profits. In case (2), add edge (s,t) to T' (it has profit at least $1/2$), and remove some edge other than (s,t) from the resulting cycle (this edge has profit at most $1/2$). Again, the resulting tree is heavier than T' in each of the d profits. $\square$

In the rest of the proof, i will range over $[d]$. The i-th profit of a spanning tree $T \in \mathcal{T}$, which we will denote by $v^{(i)}(T)$, can be written as follows

$$v^{(i)}(T) = v^{(i)}(s,t) + \sum_{j=1}^{n-2} \left(v^{(i)}(s,u_j)x_j + v^{(i)}(t,u_j)(1-x_j) \right),$$

where $x_j = x_j(T) = 1$ if edge (s,u_j) is in the tree and $x_j = 0$ otherwise, i. e., if edge (t,u_j) is in the tree. We have

$$\left(v^{(i)}(s,u_j)x_j + v^{(i)}(t,u_j)(1-x_j) \right) = \frac{v^{(i)}(s,u_j) + v^{(i)}(t,u_j)}{2} + \left(v^{(i)}(s,u_j) - v^{(i)}(t,u_j) \right) \left(x_j - \frac{1}{2} \right).$$

Now, to compute the lower bound on the expected size of the Pareto set we reveal the v's in two steps: first we reveal $\left(v^{(i)}(s,u_j) + v^{(i)}(t,u_j) \right)$ for all u_j. Then the conditional distribution of each $\left(v^{(i)}(s,u_j) - v^{(i)}(t,u_j) \right)$ is symmetric (but can be different for different i). Thus the i-th profit of $T \in \mathcal{T}$ is

$$v^{(i)}(T) = \sum_{j \in [n-2]} \left(v^{(i)}(s,u_j) - v^{(i)}(t,u_j) \right) \left(x_j - \frac{1}{2} \right) + A^{(i)},$$

where

$$A^{(i)} = v^{(i)}(s,t) + \sum_{j \in [n-2]} \frac{v^{(i)}(s,u_j) + v^{(i)}(t,u_j)}{2}.$$

Since $A^{(i)}$ is common to all trees, only the first sum in the profit matters in determining Pareto-optimality. Now we are in the situation dealt with by Corollary 2.3: for each fixing of $\left(v^{(i)}(s,u_j) + v^{(i)}(t,u_j) \right)$, we get an instance of Corollary 2.3, and thus a lower bound of

$$\left(\frac{n-3}{2(d-1)}\right)^{d-1}.$$

Since this holds for each fixing of $\left(v^{(i)}(s,u_j)+v^{(i)}(t,u_j)\right)$, we get that the same lower bound holds for the expectation without conditioning. □

4 0-1 Knapsack

We prove Theorem 1.3.

Proof of Theorem 1.3. To show our lower bound we will use the obvious one-to-one map between our basic problem with d objectives and the profits of the knapsack problem: let $v^{(1)},\ldots,v^{(d)}$ be an instance of our basic problem with all the $v_j^{(i)}$ being chosen uniformly at random from $[-1/2,1/2]$. Now the profits p are obtained from the v's in the natural way: $p_j^{(i)} = v_j^{(i)} + 1/2$. In general, the set of Pareto optima for these two problems (the basic problem instance and its corresponding knapsack instance) are not the same. We will focus instead on the better behaved set $\mathcal{S} \subseteq \{0,1\}^n$ of solutions having exactly $\lfloor n/2 \rfloor$ ones. From Corollary 2.4 we get that, in the basic problem restricted to $\mathcal{S}$, the expected number of Pareto optima is at least $\Omega_d(n^{d-1.5})$ using the well-known approximation $\binom{n}{\lfloor n/2 \rfloor} = \Theta(2^n/\sqrt{n})$.

Now we claim that if $x \in \mathcal{S}$ is Pareto-optimal in the restricted basic problem, then it is also Pareto-optimal in the corresponding (unrestricted) knapsack problem. Let $y \in \{0,1\}^n$ be different from x. There are two cases: if y has more than $\lfloor n/2 \rfloor$ ones, then it cannot dominate x, as y has a strictly higher weight (recall that all the weights are 1). If y has at most $\lfloor n/2 \rfloor$ ones, then enlarge this solution arbitrarily to a solution $y' \succeq y$ with exactly $\lfloor n/2 \rfloor$ ones. The maximality of x implies that y' is worse in some profit, and so is y, as the profits are non-negative. □

5 Lower bound for general solution sets

For proving Theorem 1.4 we consider n_0 objects with weights $1/(2n_0)$ and profits $p_j^{(i)}$ chosen uniformly at random from $[0,1/\phi]$ for $\phi \geq 2d$. Further, let $\mathcal{S}_0 \subseteq \{0,1\}^{n_0}$ be the set of solutions having exactly $\lfloor n_0/2 \rfloor$ ones. Since scaling weights and profits does not change the Pareto set, this instance is essentially the same instance as the one constructed in the proof of Theorem 1.3 and, hence, the expected number of Pareto optima is at least $\Omega_d(n_0^{d-1.5})$. Note, that all profits are chosen according to density functions bounded by ϕ and that the total weight W of these n_0 objects is $W = n_0/(2n_0) < 1$.

Now we use additional objects to clone this initial Pareto set several times to obtain a Pareto set whose size also depends on ϕ. The new objects are divided into n_1 groups each consisting of d objects, one for each profit function. The objects of group ℓ all have the same weight $w_\ell = (d+1)^{\ell-1}$ and a low profit in all but their dedicated profit function. Specifically, the i-th profit $q_{\ell j}^{(i)}$ of object j in group ℓ lies in the interval $Q_{\ell j}^{(i)}$ of length $\lceil m_\ell \rceil / \phi$, where

$$Q_{\ell j}^{(i)} = \begin{cases} \left[m_\ell - \lceil m_\ell \rceil / \phi, m_\ell\right] & \text{if } i = j, \\ \left[0, \lceil m_\ell \rceil / \phi\right] & \text{if } i \neq j. \end{cases}$$

The value m_ℓ is defined by the recurrence

$$m_\ell - \frac{m_\ell + 1}{\phi} = \frac{n_0}{\phi} + \sum_{k=1}^{\ell-1} \left(m_k + d \cdot \frac{m_k + 1}{\phi} \right) + (d-1) \cdot \frac{m_\ell + 1}{\phi} \tag{5.1}$$

and can be expressed explicitly (see Lemma A.1) by

$$m_\ell = \left(\frac{2\phi}{\phi - d} \right)^{\ell - 1} \cdot \left(\frac{n_0 + d}{\phi - d} + \frac{d}{\phi + d} \right) - \frac{d}{\phi + d}, \tag{5.2}$$

i. e., m_ℓ increases exponentially with ℓ. As can be seen from the recurrence (5.1),

$$m_\ell \geq d \cdot \frac{m_\ell + 1}{\phi} \geq \frac{\lceil m_\ell \rceil}{\phi},$$

i. e., the intervals $Q_{\ell j}^{(i)}$ are contained in the interval $[0, m_\ell]$. For the moment the profits $q_{\ell j}^{(i)} \in Q_{\ell j}^{(i)}$ are fixed values that might be larger than 1.

For $\ell \geq 1$ let $\mathcal{S}_\ell := \mathcal{S}_{\ell-1} \times \{0,1\}^d = \mathcal{S}_0 \times \{0,1\}^{\ell \cdot d}$ be the set of solutions for the instance with the n_0 initial objects and the $\ell \cdot d$ objects of the groups $1, \dots, \ell$, i. e., a solution is valid iff it uses exactly $\lfloor n_0/2 \rfloor$ of the initial objects. Assume that the Pareto set $p(\mathcal{S}_{\ell-1})$ has already been determined. If we add object j of group ℓ to each of these solutions, then we obtain a copy of $p(\mathcal{S}_{\ell-1})$ which is shifted roughly in the direction of the j-th unit vector in the weight–profits space. The next lemma shows that each shifted copy that we can obtain by additionally using a subset of the set of objects from group ℓ is part of the Pareto set $p(\mathcal{S}_\ell)$.

Lemma 5.1. *For $\ell \geq 1$, $p(\mathcal{S}_\ell) = p(\mathcal{S}_{\ell-1}) \times \{0,1\}^d$. In particular, $|p(\mathcal{S}_{n_1})| = 2^{n_1 d} \cdot |p(\mathcal{S}_0)|$.*

Proof. Consider a Pareto-optimal solution $(x,y) \in \mathcal{S}_{\ell-1} \times \{0,1\}^d$. If $x \notin p(\mathcal{S}_{\ell-1})$, then there is a partial solution $x' \in p(\mathcal{S}_{\ell-1})$ such that $x' \succ x$. By construction of $\mathcal{S}_\ell$, (x',y) is a valid solution and we obtain $(x',y) \succ (x,y)$. This contradicts the assumption that (x,y) is Pareto-optimal. Hence, $p(\mathcal{S}_\ell) \subseteq p(\mathcal{S}_{\ell-1}) \times \{0,1\}^d$.

In the remainder we show $p(\mathcal{S}_\ell) \supseteq p(\mathcal{S}_{\ell-1}) \times \{0,1\}^d$. For a vector $y \in \{0,1\}^d$ let $p^y(\mathcal{S}_\ell) = p(\mathcal{S}_{\ell-1}) \times \{y\}$. As $p^y(\mathcal{S}_\ell)$ can be interpreted as a shifted copy of the Pareto set $p(\mathcal{S}_{\ell-1})$ in the weight–profits space, vectors in this set do not dominate each other. It remains to show that solutions from different copies do not dominate each other. For this, consider solutions $(x_1,y_1) \in p^{y_1}(\mathcal{S}_\ell)$ and $(x_2,y_2) \in p^{y_2}(\mathcal{S}_\ell)$ where $y_1 \neq y_2 \in \{0,1\}^d$. If (x_1,y_1) uses more objects from group ℓ than (x_2,y_2), then the difference in weight between (x_1,y_1) and (x_2,y_2) is at least

$$\begin{aligned} w_\ell - W - \sum_{k=1}^{\ell-1} d\cdot w_k &= (d+1)^{\ell-1} - W - \sum_{k=1}^{\ell-1} d\cdot(d+1)^{k-1} \\ &= (d+1)^{\ell-1} - W - \sum_{k=1}^{\ell-1}(d+1)^k + \sum_{k=1}^{\ell-1}(d+1)^{k-1} \\ &= 1-W > 0\,. \end{aligned}$$

Hence, solution (x_1,y_1) is heavier than solution (x_2,y_2) and thus $(x_1,y_1) \not\succ (x_2,y_2)$. If (x_1,y_1) uses at most as many objects as (x_2,y_2) from group ℓ, then there is an object j of group ℓ used by (x_2,y_2), but not by (x_1,y_1). Let us analyze the j-th profit of (x_1,y_1) and (x_2,y_2). As (x_2,y_2) uses object j, its j-th profit is at least $m_\ell - \lceil m_\ell \rceil/\phi > m_\ell - (m_\ell+1)/\phi$. Since (x_1,y_1) can use at most all objects from group ℓ except object j, the j-th profit of (x_1,y_1) is bounded from above by

$$n_0\cdot\frac{1}{\phi} + \sum_{k=1}^{\ell-1}\left(m_k + (d-1)\cdot\frac{\lceil m_k\rceil}{\phi}\right) + (d-1)\cdot\frac{\lceil m_\ell\rceil}{\phi}\,,$$

which is at most $m_\ell - (m_\ell+1)/\phi$ by the recursive definition of m_ℓ (see equation (5.1)). Consequently, (x_2,y_2) has a higher j-th profit than (x_1,y_1) and, thus, $(x_1,y_1) \not\succ (x_2,y_2)$. This yields $p(\mathcal{S}_\ell) \supseteq p(\mathcal{S}_{\ell-1}) \times \{0,1\}^d$.

The second part of the claim follows from $|p(\mathcal{S}_{n_1})| = |p(\mathcal{S}_0)\times\{0,1\}^{n_1 d}| = 2^{n_1 d}\cdot|p(\mathcal{S}_0)|$. □

As mentioned earlier the profits $q_{\ell j}^{(i)}$ might be larger than 1. We resolve this problem by splitting the objects into $k_\ell = \lceil m_\ell \rceil$ objects with weights w_ℓ/k_ℓ and profits $q_{\ell j}^{(i)}/k_\ell$ lying in the interval $Q_{\ell j}^{(i)}/k_\ell$. By the choice of the intervals $Q_{\ell j}^{(i)}$, these new intervals have length $1/\phi$. Furthermore, each interval $Q_{\ell j}^{(i)}/k_\ell$ is a subset of $[0,m_\ell] \subseteq [0,k_\ell]$. Hence, the scaled intervals are subsets of $[0,1]$. In addition we use a set $\mathcal{S} \subseteq \mathcal{S}_0 \times \prod_{\ell=1}^{n_1}\{0,1\}^{k_\ell\cdot d}$ consisting of all solutions that use all k_ℓ parts of one object of group ℓ or no single one. This simulates using the whole object j or not using it at all. In this way we do not change the size of the Pareto set, but we use more objects. Recall that we do not consider the standard knapsack problem where each object can be used independently of the other objects. In our variant of the knapsack problem we can specify the set $\mathcal{S}$ of feasible combinations of objects. Since Lemma 5.1 holds for all profits $q_{\ell j}^{(i)} \in Q_{\ell j}^{(i)}$, we obtain

Corollary 5.2. *The size of the Pareto set of the 0-1 knapsack instance constructed above is*

$$\Omega_d\left(2^{n_1 d} n_0^{d-1.5}\right).$$

The bound in Corollary 5.2 is still expressed in the numbers n_0 and n_1 and not in the total number N of objects used. This number can be bounded as follows.

Lemma 5.3. *The number N of objects is bounded by*

$$N \le n_0 + d\cdot n_1 + d\cdot\frac{n_0+2d}{\phi+d}\cdot\left(\frac{2\phi}{\phi-d}\right)^{n_1}.$$

Proof. Recall that our instance consists of n_0 initial objects and n_1 groups $\ell = 1, \dots, n_1$ containing d objects each of which we split into k_ℓ smaller objects. Hence, the claim holds true since

$$N = n_0 + \sum_{\ell=1}^{n_1} d \cdot k_\ell \le n_0 + \sum_{\ell=1}^{n_1} d \cdot (m_\ell + 1) = n_0 + d \cdot n_1 + d \cdot \sum_{\ell=1}^{n_1} m_\ell$$

and

$$\sum_{\ell=1}^{n_1} m_\ell \le \left(\frac{n_0+d}{\phi-d} + \frac{d}{\phi+d}\right) \cdot \sum_{\ell=1}^{n_1} \left(\frac{2\phi}{\phi-d}\right)^{\ell-1} \le \left(\frac{n_0+d}{\phi-d} + \frac{d}{\phi-d}\right) \cdot \frac{\left(\frac{2\phi}{\phi-d}\right)^{n_1}}{\frac{2\phi}{\phi-d} - 1}$$
$$= \frac{n_0+2d}{\phi+d} \cdot \left(\frac{2\phi}{\phi-d}\right)^{n_1},$$

where the first inequality is due to the explicit formula (5.2) of m_ℓ. □

Now we can prove Theorem 1.4.

Proof of Theorem 1.4. Using the instance described above, our goal is to choose the parameters n_0 and n_1 in such a way that the expected number of Pareto-optimal solutions becomes large upon condition that the number of required objects is at most n. We only consider sufficiently large values of n and ϕ. In particular, $n \ge 8d$ and $\phi \ge 2d$. This is not a restriction as d is constant. For the moment let us assume that $\phi/d \le 2^{\frac{n}{4d}-1}$. This is the interesting case leading to the first term in the minimum in Theorem 1.4. Let $n_0 = \lfloor \hat{n}_0 \rfloor$ and $n_1 = \lfloor \hat{n}_1 \rfloor$, where

$$\hat{n}_1 = \log_2\left(\frac{\phi}{d}\right) \in \left[1, \frac{n-4d}{4d}\right],$$
$$\hat{n}_0 = \frac{n - 2d \cdot \left(\frac{\phi}{d}\right)^h - d \cdot \hat{n}_1}{1 + \left(\frac{\phi}{d}\right)^h}, \qquad \text{and}$$
$$h = h(\phi, d) = \log_2\left(\frac{2\phi}{\phi-d}\right) - 1 = \log_2\left(\frac{\phi}{\phi-d}\right).$$

In Lemma A.2 we show that $(\phi/d)^h$ takes only values in $[1,2]$ provided $\phi > d$. This implies

$$\hat{n}_0 \ge \frac{n - 4d - (n-4d)/4)}{3} = \frac{n-4d}{4} \ge d$$

since $n \ge 8d$ and, thus, $n_0, n_1 \ge 1$. By Lemma 5.3 the number N of objects of this instance is bounded by

$$N \le n_0 + d \cdot n_1 + d \cdot \frac{n_0+2d}{\phi+d} \cdot \left(\frac{2\phi}{\phi-d}\right)^{n_1} \le \hat{n}_0 + d \cdot \hat{n}_1 + d \cdot \frac{\hat{n}_0+2d}{\phi} \cdot \left(\frac{2\phi}{\phi-d}\right)^{\hat{n}_1}$$
$$= \hat{n}_0 + d \cdot \hat{n}_1 + (\hat{n}_0 + 2d) \cdot \frac{d}{\phi} \cdot \left(\frac{\phi}{d}\right)^{\log_2\left(\frac{2\phi}{\phi-d}\right)} = \hat{n}_0 + d \cdot \hat{n}_1 + (\hat{n}_0 + 2d) \cdot \left(\frac{\phi}{d}\right)^h$$
$$= \hat{n}_0 \cdot \left(1 + \left(\frac{\phi}{d}\right)^h\right) + 2d \cdot \left(\frac{\phi}{d}\right)^h + d \cdot \hat{n}_1 = n$$

by definition of $\hat{n}_0$, that is, the number N of objects we actually use is at most n, as required. Due to Corollary 5.2 the expected number of Pareto-optimal solutions of our instance is

$$\Omega_d\left(2^{n_1 d} \cdot n_0^{d-1.5}\right) = \Omega_d\left(2^{\hat{n}_1 d} \cdot \hat{n}_0^{d-1.5}\right) = \Omega_d\left(\left(\frac{\phi}{d}\right)^d \cdot \left(\frac{n}{4} - d\right)^{d-1.5}\right) = \Omega_d\left(n^{d-1.5}\phi^d\right).$$

In the case $\phi/d > 2^{\frac{n}{4d}-1}$ we construct the same instance as above, but for a maximum density $\phi' = d \cdot 2^{\frac{n}{4d}-1} \in [2d, \phi)$. Consequently, $\hat{n}_1 = n/(4d) - 1$. As above, the expected size of the Pareto set is $\Omega_d\left(n^{d-1.5} \cdot 2^{\frac{n}{4}-d}\right) = 2^{\Theta_d(n)}$. □

Note that for $d = 1$ Beier and Vöcking proved a lower bound of $\Omega(n^2)$ for the expected number of Pareto-optimal solutions in the smoothed model using profits chosen uniformly at random from $[0,1]$ and exponentially increasing weights [2]. We can apply the same construction as for general values d using n_0 objects of this instance as initial objects and obtain:

Corollary 5.4. *There is a ϕ-smooth instance for the knapsack problem with one profit function where the expected number of Pareto-optimal solutions is* $\min\{\Omega(n^2\phi), 2^{\Theta(n)}\}$.

The bound of Corollary 5.4 is asymptotically tight due to the upper bound of $O(n^2\phi)$ proved in [1].

6 Discussion and conclusion

We proved lower bounds for the average and smoothed number of Pareto optima by introducing geometric arguments to this setting. Our lower bound for d random objectives is of the form $\Omega_d(n^{d-1})$. The best upper bound we know, even for $\phi = 1$, is that of Moitra and O'Donnell [15] which is of the form $O_d(n^{2d-2})$, ignoring the dependence on ϕ. Thus there is a gap between the upper and lower bounds. As mentioned before, the number of Pareto optima for the case when 2^n points are chosen uniformly at random from $[-1,1]^d$ is $\Theta_d(n^{d-1})$. For the case with one adversarial and d quasiconcave ϕ-smooth objectives we proved a lower bound of $\Omega_d(n^{d-1.5}\phi^d)$. This bound matches the best known upper bound of $O_d(n^{2d}\phi^d)$ due to Brunsch and Röglin [7] with regard to the dependence on ϕ but there is still a gap when considering the dependence on n.

Do lower bounds similar to ours hold for any sufficiently large feasible set $\mathcal{S}$? Our techniques can show this for natural objectives, but require arguments tailored to the specific objective. It is desirable to have a general lower bound technique that works for all sufficiently large $\mathcal{S}$. Also, in smoothed lower bounds, to get a good dependence on ϕ our technique requires a very special choice of $\mathcal{S}$. So, a more general question is whether we can prove lower bounds with strong dependence on ϕ for all sufficiently large $\mathcal{S}$.

We now briefly discuss some difficulties in proving lower bounds for general $\mathcal{S}$. One approach to this end is to show a lower bound on the expected size of the Pareto set that depends only on $|\mathcal{S}|$, n and d. Our general technique was to first reduce the problem to giving a lower bound on the expected number of vertices in the projection of the convex hull of the points in $\mathcal{S}$ to a random subspace of dimension d. A special distribution which is instructive to consider here, and also interesting in its own right, is given by the case when we project to a d-dimensional space chosen uniformly at random. The expected number of vertices in the projection has been studied for the special cases of the simplex, the cube, and the cross-polytope (see Schneider [20]). But understanding this number for arbitrary 0/1-polytopes seems difficult. When the subspace to be projected to is of dimension $n-1$, we can write the expected number of vertices in the projection as $C \cdot \sum_{v \in V} a(v)$, where $a(v)$ is the solid angle of the cone polar to the tangent cone at vertex v, and C is a constant depending on n. (Suitable generalizations of this formula are easy to obtain for projection to dimensions smaller than $n-1$, but the case of dimension $n-1$ is sufficient for our purpose here.) This captures the intuitive fact that if the polytope is very pointy at vertex v, then v is more likely to be a vertex in the convex hull. It is natural to ask: given k, what is the $\mathcal{S} \subseteq \{0,1\}^n$ with $|\mathcal{S}| = k$ that minimizes this expectation? Intuitively, the sum of angles $a(v)$ could be minimized when the vertices are close together, as in a Hamming ball. Note the high-level similarity of the problem at hand to the vertex-isoperimetric inequality for the Boolean cube (see, e. g., Theorem 5, Chapter 16 of [5]). Unfortunately, our numerical experiments show that Hamming balls are not the minimizers of the expected number of vertices of a random projection. In particular, there exists a subset of the vertices of the 5-dimensional Boolean cube, of size 16, with smaller expected number of vertices in the projection to a uniformly random two-dimensional plane compared to the Hamming ball with 16 vertices.

7 Acknowledgements

We thank Yusu Wang for helpful discussions. We are also very thankful of the Isaac Newton Institute for Mathematical Sciences, Cambridge, UK, where part of this work was carried out during the programme on Discrete Analysis. We also thank the anonymous referees of this paper for detailed comments that helped improve the presentation.

A Technical proofs

Lemma A.1. *The explicit formula of m_ℓ given by the recurrence*

$$m_\ell - \frac{m_\ell + 1}{\phi} = \frac{n_0}{\phi} + \sum_{k=1}^{\ell-1} \left(m_k + d \cdot \frac{m_k + 1}{\phi} \right) + (d-1) \cdot \frac{m_\ell + 1}{\phi}$$

is

$$m_\ell = \left(\frac{2\phi}{\phi - d} \right)^{\ell-1} \cdot \left(\frac{n_0 + d}{\phi - d} + \frac{d}{\phi + d} \right) - \frac{d}{\phi + d}\,.$$

Proof. The recurrence formula is equivalent to the formula

$$m_\ell - d \cdot \frac{m_\ell + 1}{\phi} = \frac{n_0}{\phi} + \sum_{k=1}^{\ell-1} \left(m_k + d \cdot \frac{m_k + 1}{\phi} \right),$$

which yields $m_1 = (n_0 + d)/(\phi - d)$ and

$$\frac{\phi - d}{\phi} \cdot m_\ell - \frac{d}{\phi} = \frac{\phi - d}{\phi} \cdot m_{\ell-1} - \frac{d}{\phi} + m_{\ell-1} + d \cdot \frac{m_{\ell-1} + 1}{\phi} = 2m_{\ell-1}$$

for $\ell \geq 2$. With $\alpha = 2\phi/(\phi - d)$ and $\beta = d/(\phi - d)$, we obtain

$$\begin{aligned} m_\ell &= \alpha \cdot m_{\ell-1} + \beta = \alpha^{\ell-1} \cdot m_1 + \left(\sum_{k=0}^{\ell-2} \alpha^k\right) \cdot \beta = \alpha^{\ell-1} \cdot m_1 + \frac{\alpha^{\ell-1} - 1}{\alpha - 1} \cdot \beta \\ &= \alpha^{\ell-1} \cdot \left(m_1 + \frac{\beta}{\alpha - 1}\right) - \frac{\beta}{\alpha - 1} = \left(\frac{2\phi}{\phi - d}\right)^{\ell-1} \cdot \left(\frac{n_0 + d}{\phi - d} + \frac{d}{\phi + d}\right) - \frac{d}{\phi + d}. \end{aligned}$$

□

Lemma A.2. *Let $h = h(\phi, d) = \log_2\left(\frac{\phi}{\phi - d}\right)$. Then, $1 \leq (\phi/d)^h \leq 2$ for any $\phi > d$.*

Proof. With $x = \phi/d - 1 > 0$ we can restate the inequalities as

$$1 \leq (1+x)^{\log_2(1+1/x)} \leq 2$$

which is equivalent to $0 \leq \hat{f}(x) \leq 1$ for $\hat{f}(x) = \log_2(1+x) \cdot \log_2(1+1/x)$. Obviously, the first inequality is true. In the remainder of this proof we show that $\hat{f}(x)$ is maximal for $x = 1$ (and, consequently, $\hat{f}(x) \leq \hat{f}(1) = 1$ for all $x > 0$) which is equivalent to showing that $f(x) = \ln(1+x) \cdot \ln(1+1/x)$ is maximal for $x = 1$. Since $f(x) = f(1/x)$ we focus on values $x \geq 1$.

The derivative of f is

$$f'(x) = \frac{1}{1+x} \cdot \ln(1+1/x) + \ln(1+x) \cdot \frac{1}{1+1/x} \cdot \left(-\frac{1}{x^2}\right) = \frac{g(x)}{x \cdot (1+x)}$$

for $g(x) = x \cdot \ln(1+1/x) - \ln(1+x)$. We show that $f'(x) < 0$ for $x > 1$ by showing that g is monotonically decreasing for $x > 1$. This is sufficient due to $g(1) = 0$. The derivative of g is

$$\begin{aligned} g'(x) &= \ln(1+1/x) + x \cdot \frac{1}{1+1/x} \cdot (-1/x^2) - 1/(1+x) \\ &= \ln(1+1/x) - 2/(1+x) \leq 1/x - 2/(1+x) = (1-x)/(x(1+x)) < 0 \end{aligned}$$

for $x > 1$. The first inequality stems from the inequality $\ln(1+x) \leq x$ for all $x > 0$. □

References

[1] RENÉ BEIER, HEIKO RÖGLIN, AND BERTHOLD VÖCKING: The smoothed number of Pareto optimal solutions in bicriteria integer optimization. In *Proc. 12th Ann. Conf. on Integer Programming and Combinatorial Optimization (IPCO'07)*, pp. 53–67. Springer, 2007. [doi:10.1007/978-3-540-72792-7_5] 239, 251

[2] RENÉ BEIER AND BERTHOLD VÖCKING: Random knapsack in expected polynomial time. *J. Comput. Syst. Sci.*, 69(3):306–329, 2004. Preliminary version in STOC'03. [doi:10.1016/j.jcss.2004.04.004] 239, 240, 251

[3] RENÉ BEIER AND BERTHOLD VÖCKING: Typical properties of winners and losers in discrete optimization. *SIAM J. Comput.*, 35(4):855–881, 2006. Preliminary version in STOC'04. [doi:10.1137/S0097539705447268] 239

[4] JON LOUIS BENTLEY, HSIANG-TSUNG KUNG, MARIO SCHKOLNICK, AND CLARK D. THOMPSON: On the average number of maxima in a set of vectors and applications. *J. ACM*, 25(4):536–543, 1978. [doi:10.1145/322092.322095] 241, 242, 243

[5] BÉLA BOLLOBÁS: *Combinatorics: Set systems, hypergraphs, families of vectors and combinatorial probability*. Cambridge University Press, Cambridge, 1986. 252

[6] TOBIAS BRUNSCH AND HEIKO RÖGLIN: Lower bounds for the smoothed number of Pareto optimal solutions. In *Theory and Applications of Models of Computation - 8th Annual Conference (TAMC'11)*, LNCS 6648, pp. 416–427. Springer, 2011. [doi:10.1007/978-3-642-20877-5_41, arXiv:abs/1012.1163] 237

[7] TOBIAS BRUNSCH AND HEIKO RÖGLIN: Improved smoothed analysis of multiobjective optimization. In *Proc. 44th STOC*, pp. 407–426. ACM Press, 2012. [doi:10.1145/2213977.2214016, arXiv:abs/1111.1546] 240, 242, 251

[8] CHRISTIAN BUCHTA: On the average number of maxima in a set of vectors. *Inf. Process. Lett.*, 33(2):63–65, 1989. [doi:10.1016/0020-0190(89)90156-7] 241

[9] LUC DEVROYE: A note on finding convex hulls via maximal vectors. *Inf. Process. Lett.*, 11(1):53–56, 1980. [doi:10.1016/0020-0190(80)90036-8] 241

[10] DAVID L. DONOHO AND JARED TANNER: Counting the faces of randomly-projected hypercubes and orthants, with applications. *Discrete Comput. Geom.*, 43(3):522–541, 2010. [doi:10.1007/s00454-009-9221-z, arXiv:0807.3590] 243

[11] MATTHIAS EHRGOTT: *Multicriteria Optimization (2nd ed.)*. Springer, 2005. [doi:10.1007/3-540-27659-9] 238, 240

[12] JACOB E. GOODMAN AND JOSEPH O'ROURKE: *Handbook of Discrete and Computational Geometry*. Discrete Mathematics and its Applications. Chapman & Hall/CRC, 2004. 243

[13] NAVIN GOYAL AND LUIS RADEMACHER: Lower bounds for the average and smoothed number of Pareto optima. In *IARCS Annual Conf. on Foundations of Software Technology and Theoretical Computer Science (FSTTCS'12)*, pp. 58–69. Springer, 2012. [doi:10.4230/LIPIcs.FSTTCS.2012.58, arXiv:abs/1107.3876] 237

[14] FABRIZIO GRANDONI, RAMAMOORTHI RAVI, AND MOHIT SINGH: Iterative rounding for multiobjective optimization problems. In *Proc. 17th Ann. European Symp. on Algorithms (ESA'09)*, pp. 95–106, 2009. [doi:10.1007/978-3-642-04128-0_9] 238

[15] ANKUR MOITRA AND RYAN O'DONNELL: Pareto optimal solutions for smoothed analysts. *SIAM J. Comput.*, 41(5):1266–1284, 2012. Preliminary version in STOC'11. [doi:10.1137/110851833] 238, 240, 242, 251

[16] MATTHIAS MÜLLER-HANNEMANN AND KARSTEN WEIHE: Pareto shortest paths is often feasible in practice. In *Algorithm Engineering, 5th Internat. Workshop (WAE'01)*, pp. 185–197. Springer, 2001. [doi:10.1007/3-540-44688-5_15] 238

[17] GEORGE L. NEMHAUSER AND ZEV ULLMANN: Discrete dynamic programming and capital allocation. *Management Science*, 15(9):494–505, 1969. [doi:10.1287/mnsc.15.9.494] 238

[18] FRANCO P. PREPARATA AND MICHAEL IAN SHAMOS: *Computational Geometry: An Introduction*. Texts and Monographs in Computer Science. Springer, New York, 1985. [doi:10.1007/978-1-4612-1098-6] 242

[19] HEIKO RÖGLIN AND SHANG-HUA TENG: Smoothed analysis of multiobjective optimization. In *Proc. 50th FOCS*, pp. 681–690. IEEE Comp. Soc. Press, 2009. [doi:10.1109/FOCS.2009.21] 238, 240

[20] ROLF SCHNEIDER: Recent results on random polytopes: A survey. *Bollettino dell'Unione Matematica Italiana*, Ser. (9), 1:17–40, 2008. 252

[21] ROLF SCHNEIDER AND WOLFGANG WEIL: *Stochastic and integral geometry*. Probability and its Applications (New York). Springer, Berlin, 2008. [doi:10.1007/978-3-540-78859-1] 243

[22] DANIEL A. SPIELMAN AND SHANG-HUA TENG: Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. *J. ACM*, 51(3):385–463, 2004. Preliminary version in STOC'01. [doi:10.1145/990308.990310] 238, 239

[23] J.G. WENDEL: A problem in geometric probability. *Mathematica Scandinavica*, 11:109–112, 1962. Found at Mathematica Scandinavica. 243

AUTHORS

Tobias Brunsch
Ph. D. student
University of Bonn
brunsch@cs.uni-bonn.de

Navin Goyal
researcher
Microsoft Research India
navingo@microsoft.com
`http://research.microsoft.com/en-us/people/navingo/`

Luis Rademacher
assistant professor
Ohio State University
lrademac@cse.ohio-state.edu
`http://www.cse.ohio-state.edu/~lrademac/`

Heiko Röglin
professor
University of Bonn
roeglin@cs.uni-bonn.de
http://www.roeglin.org

ABOUT THE AUTHORS

TOBIAS BRUNSCH studied Computer Science and Mathematics at the Chemnitz University of Technology. In 2010, he joined the group of Heiko Röglin as a Ph. D. student. His research focuses on probabilistic analysis of algorithms.

NAVIN GOYAL received his Ph. D. from Rutgers University in 2005. After two postdoctoral stints at McGill University and Georgia Institute of Technology he returned to his native India in 2009 where he now works at Microsoft Research. His research interests are in theoretical computer science.

LUIS RADEMACHER graduated with a Ph. D. in mathematics from the Massachusetts Institute of Technology in 2007 under the guidance of Santosh Vempala. He spent two years as a Postdoctoral Fellow in the College of Computing at the Georgia Institute of Technology. He joined the Computer Science and Engineering Department at The Ohio State University in 2009. His interest lie around computational learning theory and random structures and algorithms.

HEIKO RÖGLIN graduated from RWTH Aachen University in 2008 under the supervision of Berthold Vöcking. He spent a year as a postdoc with Shang-Hua Teng at Boston University and a year as an assistant professor at Maastricht University before joining the University of Bonn in 2010. His research focuses on probabilistic analysis of algorithms.