

Lower Bounds for the Average and Smoothed Number of Pareto Optima

Navin Goyal¹ and Luis Rademacher²

1 Microsoft Research India
navingo@microsoft.com

2 Computer Science and Engineering
Ohio State University
lrademac@cse.ohio-state.edu

Abstract

Smoothed analysis of multiobjective 0–1 linear optimization has drawn considerable attention recently. In this literature, 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 is the central object of study. In this paper, 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 lower bounding the number of Pareto optima to results in discrete geometry and geometric probability connected to arrangements of hyperplanes. We use our basic result to derive (1) To our knowledge, the first lower bound for natural multiobjective optimization problems. We illustrate this for the maximum spanning tree problem with randomly chosen edge weights. Our technique is sufficiently flexible to yield such lower bounds for other standard objective functions studied in this setting (such as multiobjective shortest path, TSP tour, matching). (2) Smoothed lower bound of $\min\{\Omega_d(n^{d-1.5}\phi^{(d-\log d)(1-\Theta(1/\phi))}), 2^{\Theta(n)}\}$ for the 0–1 knapsack problem with d profits for ϕ -semirandom distributions for a version of the knapsack problem. This improves the recent lower bound of Brunsch and Röglin.

1998 ACM Subject Classification F.2.2 Geometrical problems and computations

Keywords and phrases geometric probability, smoothed analysis, multiobjective optimization, Pareto optimal, knapsack

Digital Object Identifier 10.4230/LIPIcs.FSTTCS.2012.58

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 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., [12, 13, 17] for references. It is the latter approach using Pareto-optimality that concerns us in this paper. A Pareto-optimal solution is a solution with the property that no other set 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)



© John Q. Open and Joan R. Access;

licensed under Creative Commons License NC-ND

32nd Int'l Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2012).

Editors: D. D'Souza, J. Radhakrishnan, and K. Telikepalli; pp. 58–69



Leibniz International Proceedings in Informatics

LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

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., [14]). However, in many cases of interest, the Pareto set is typically not too large. Thus, if the Pareto set is small and can be generated efficiently, then it can be treated possibly with human assistance to choose among the few alternatives. Pareto sets are also used in heuristics for optimization problems (e.g. [15]). To explain why Pareto sets are frequently small in practice, multiobjective optimization has recently been studied from the view-point of smoothed analysis [20]. We introduce some notation before describing this work.

Notation.

For positive integer n , we denote the set $\{1, 2, \dots, n\}$ by $[n]$. 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 $v^{(i)} : \mathcal{S} \rightarrow \mathbb{R}$, given by $v^{(i)}(x) = \sum_{j \in [n]} v_j^{(i)} x_j$, 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 is the following.

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

For the special case of the multiobjective 0–1 knapsack problem we have $d + 1$ objectives: One of the objectives will be the weight $w = (w_1, \dots, w_n)$, which should be minimized, and the other d objectives will be the profits as before: $v^{(i)} = (v_1^{(i)}, \dots, v_n^{(i)})$ for $i \in [d]$, and all the entries in w and $v^{(i)}$ come from $[0, 1]$.

Let V be the $d \times n$ matrix with rows $v^{(1)}, \dots, v^{(d)}$. 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 of b dominating a by $b \succ a$. 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 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 a set of points X in Euclidean space, let $p(X)$ denote the number of Pareto-optima in X .

Smoothed analysis.

For our multiobjective optimization problem (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 analysis of algorithms introduced by Spielman and Teng [20] 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 used for a variety of algorithms. Beier and Vöcking [3] studied bicriteria 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 stronger notion of smoothed analysis. In one version of their model, each entry of the matrix V is an independent random variable taking values in $[0, 1]$ with the restriction

that each has probability density function bounded 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 -semirandom distributions. For more generality, one of the rows of V could be chosen fully adversarially (deterministically). As ϕ is increased, the semirandom model becomes more like the worst case model. With the exception of Theorem 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 bi-criteria version of the 0–1 knapsack problem with adversarial weights the expected number of Pareto optima is $O(\phi n^4)$. The result generalizes to other bicriteria optimization problems. [1] makes this generalization explicit and improves the upper bound to $O(\phi n^2)$. Röglin and Teng [17] studied the multiobjective optimization problem in the above framework. They showed that the expected size of the Pareto set with d objectives is of the form $O((\phi n)^{2^{d-2}(d+1)!})$. Moitra and O’Donnell [13] improved this upper bound to $2 \cdot (4\phi d)^{d(d+1)/2} \cdot n^{2d}$. (Again, these results allow one of the objectives to be chosen adversarially.) Very recently, this has been improved to $O(n^{2d}\phi^d)$ [6] for the mildly restricted class of quasiconcave ϕ -smooth instances. These papers [17, 13] raised the question of a lower bound on the expected number of Pareto optima.

An early average-case lower bound of $\Omega(n^2)$ was proven in [2] for the knapsack problem with a single profit vector. Their result however required an adversarial choice of exponentially increasing weights. Recently, Brunsch and Röglin [5] proved lower bounds of the form

$$\Omega_d(\min\{(n\phi)^{(d-\log_2 d) \cdot (1-\Theta(1/\phi))}, 2^{\Theta(n)}\}),$$

where Ω_d means that the constant in the asymptotic notation may depend on d . Unfortunately, the instances constructed by them use \mathcal{S} that does not seem to correspond to natural optimization problems.

Our results.

In this paper we prove lower bounds on the expected number of Pareto optima. Our basic result deals with the case when every entry in the matrix V is chosen independently from a distribution with density symmetric around the origin. Note that we do not require that the distributions be identical: Each entry can have a different distribution but we require that the distributions have a density. This generality will in fact be useful in our lower bound for the maximum spanning tree problem, Theorem 2. Note also that all entries of V are random unlike the results discussed above where one of the objectives is chosen adversarially. This makes our lower bound stronger.

► **Theorem 1 (Basic theorem).** *Suppose that each entry of a $d \times n$ random matrix V is chosen independently according to (not necessarily identical) symmetric distributions with a density. Let X denote the random set $\{Vr : r \in \{0, 1\}^n\}$. Then*

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

This implies the simpler bound $\mathbb{E}_V p(X) \geq \left(\frac{n-1}{2^{d-1}}\right)^{d-1}$.

We give two proofs of this result. The two proofs have a similar essence, but a somewhat 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 on the expected number of vertices of projections of hypercubes, as the random set in Theorem 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 the counting of Pareto-optima to lower bounding the probability that the convex hull of n random points contains the origin. This probability is known exactly by a theorem of Wendel (Theorem 6).

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 better understood.

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 on 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 tour, 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 set here lives in $\{0, 1\}^{\binom{n}{2}}$ and not in $\{0, 1\}^n$.

► **Theorem 2.** *In the d objective maximum spanning tree problem on K_n there exists a choice of 4-semirandom 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, namely the ability to choose different symmetric distributions.

In our basic theorem above, Theorem 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 multiobjective 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 $v^{(i)}$ 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 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 $\Omega_d(n^{d-1.5})$.*

Theorems 1 or 3 (depending on whether one wants a bound for non-negative or unrestricted weights and profits) can be used in a simple way as the base case of the argument with

$d + 1$ objectives in [5, Section 3] to give the following improved lower bound on the expected number of Pareto optima when the profits are ϕ -semirandom (actually, uniform in carefully chosen intervals of length at least $1/\phi$):

► **Theorem 4.** *For any fixed $d \geq 2$ (so that the constants in asymptotic notation may depend on d) and for $n \in \mathbb{N}$ and $\phi > 1$ 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 $(d + 1)$ -dimensional knapsack problem with solution set \mathcal{S} is at least

$$\min\{\Omega_d(n^{d-1.5}\phi^{(d-\log d)(1-\Theta(1/\phi))}), 2^{\Theta(n)}\}.$$

For general multiobjective optimization (basically without the restriction of entries being non-negative) the exponent of n becomes exactly d .

The technique of [5] requires \mathcal{S} to be chosen adversarially, and so this is the case for Theorem 4 above as well. To our knowledge, no non-trivial lower bounds were known before our work for natural choices of \mathcal{S} . This is addressed by our Theorems 1 ($\mathcal{S} = \{0, 1\}^n$) and 2 (\mathcal{S} is the set of spanning trees of the complete graph) above, though these Theorems are for a small constant value of ϕ , and therefore do not clarify what the dependence of ϕ should be.

Recently, Brunsch and Röglin improved the induction step of their lower bound [7]. Combining their improved result with our result yields the lower bound of $\min\{\Omega_d(n^{d-1.5}\phi^d), 2^{\Theta(n)}\}$. This lower bound should be contrasted with the aforementioned upper bounds $2 \cdot (4\phi d)^{d(d+1)/2} \cdot n^{2d}$ [13] and $O(n^{2d}\phi^d)$ [6] (this latter bound is only for the mildly restricted but natural class of quasiconcave ϕ -smooth instances; our lower bound also holds for such quasiconcave perturbations).

2 The basic theorem

In this section we prove Theorem 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 paper we found the ideas and known results that could be combined to get a more geometric proof.

2.1 First proof

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

► **Theorem 5** ([4], [16, Theorem 4.7]). *Let P be a finite subset of \mathbb{R}^d . A vertex of the convex hull of P is maximal under \preceq in at least one of the 2^d assignments of d signs $+$ and $-$ to each of the coordinates of the points of P .*

First proof of Theorem 1. The convex hull of X is a random polytope, a zonotope actually, that is, a linear image of a hypercube or, equivalently, a Minkowski sum of segments. By Theorem 5, every vertex is maximal under our partial order \preceq for at least one of the 2^d reflections involving coordinate hyperplanes. That is

$$|\text{vertices of conv } X| \leq \sum_{\epsilon \in \{-1, 1\}^d} p(X \text{ with coordinates of points flipped by signs in } \epsilon) \quad (3)$$

Our symmetry assumption followed by (3) implies¹

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

It is known [10, Theorem 1.8] that for V with columns in general position (that is, any d columns are linearly independent, which happens almost surely in our case) the number of vertices is equal to the maximal number of vertices of a d -dimensional zonotope formed as the sum of n segments, and this number is known exactly [11, 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 $\epsilon \in \{-1, 1\}^d$, the orthant associated with ϵ is $\{(\epsilon_1 x_1, \dots, \epsilon_d x_d) : (x_1, \dots, x_d) \in \mathbb{R}_+^d\}$. In particular, if ϵ is the all 1's vector then we call its associated orthant the positive orthant, and if ϵ 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 $\text{cone}(P) = \{\sum_{i=1}^k \alpha_i p_i : \alpha_i \geq 0\}$ (note that the conic hull is always convex).

We will use the following result by Wendel:

► **Theorem 6** ([21], [19, Theorem 8.2.1]). *If X_1, \dots, X_n are independent random points in \mathbb{R}^d whose distributions are symmetric with respect to 0 and such that with probability 1 all subsets of size d are linearly independent, then*

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

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 i.i.d. at random from the unit sphere.

Second proof of Theorem 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 , so we can write $\mathbb{E} p(X) = 2^n \Pr[V1 \text{ maximal}]$.

For the rest of the proof we will focus on finding a lower bound on this last probability. To understand this probability we first rewrite the event $[V1 \text{ maximal}]$ in terms of a different

¹ This idea is from [4]. It is used there in the opposite direction, that is, to get upper bounds on the expected number of vertices from upper bounds on the expected number of maximal points.

event via easy intermediate steps:

$$\begin{aligned} [V1 \text{ maximal}] &= [Vr \not\prec V1, \forall r \in \{0, 1\}^n] \\ &= [0 \not\prec V(1-r), \forall r \in \{0, 1\}^n] \\ &= [0 \not\prec Vr, \forall r \in \{0, 1\}^n]. \end{aligned}$$

Now we have $\Pr[0 \not\prec Vr, \forall r \in \{0, 1\}^n] \geq \Pr[0 \not\prec Vr, \forall r \in [0, 1]^n]$.

Event $[0 \not\prec Vr, \forall r \in [0, 1]^n]$ is the same as the event $[\text{cone}(v_1, \dots, v_n) \cap \mathbb{R}_-^d = \{0\}]$. 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.

By the separability property of convex sets (Hahn-Banach theorem) we have that there exists a hyperplane $H = \{x \in \mathbb{R}^d : \langle u, x \rangle = 0\}$ separating $\text{cone}(v_1, \dots, v_n)$ and \mathbb{R}_-^d . That is, there exists $u \in \mathbb{R}_+^d \setminus \{0\}$ such that $\text{cone}(v_1, \dots, v_n) \cdot u \geq 0$ and this implies

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

Now

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

Clearly, we have

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

Theorem 6 and the fact that the distribution of v_i is centrally symmetric and assigns measure zero to every hyperplane through 0 imply

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

We conclude:

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

◀

The following easy corollary of Theorem 1 will be useful later.

► **Corollary 7.** Theorem 1 holds also when the feasible set $\mathcal{S} = \{-1, 1\}^n$.

► **Corollary 8.** Under the assumptions of Theorem 1 and when the set of feasible solutions is $\mathcal{S} \subseteq \{0, 1\}^n$ we have

$$\mathbb{E}_V p(X) \geq \frac{|\mathcal{S}|}{2^{n+d-1}} \sum_{k=0}^{d-1} \binom{n-1}{k}. \quad (4)$$

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 distribution this probability is independent of r and by Theorem 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. ◀

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 paper. We illustrate this for the case of 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 instance of the basic problem into the instance of the problem at hand. The proof requires the full power of Theorem 1. It is worth noting that the direct use of Cor. 8 does not provide any useful bound for the case of spanning trees. The proof below is easily modified so that all profits are chosen from intervals of non-negative numbers.

We now prove Theorem 2.

Proof. 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 for convenience we denote by $\{s, t, u_1, u_2, \dots, u_{n-2}\}$. 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 $[1/2, 1]$. And for all other edges it is uniform over $[-1/2, 1/2]$. Let \mathcal{T} denote the set of spanning trees which 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 9.** For any choices of profits from the intervals as specified above, if a tree T is Pareto-optimal then $T \in \mathcal{T}$.

Proof. Fix any choice of profits as above. Suppose that a tree T' is Pareto-optimal but $T' \notin \mathcal{T}$. Then (1) either 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 $E(G)$, and then complete the remaining disconnected graph to a spanning tree using edges from $E(G)$. Clearly, the resulting tree is heavier than T' in each of the d weights. In case (2), add edge (s, t) to T' , and from the resulting cycle remove some edge other than (s, t) . Again, the resulting tree is heavier than T' in each of the d weights. ◀

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

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

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

$$\begin{aligned} & (v^{(i)}(s, u_j)x_j + v^{(i)}(t, u_j)(1 - x_j)) \\ &= \frac{v^{(i)}(s, u_j) + v^{(i)}(t, u_j)}{2} + (v^{(i)}(s, u_j) - v^{(i)}(t, u_j))(x_j - \frac{1}{2}). \end{aligned}$$

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 $(v^{(i)}(s, u_j) + v^{(i)}(t, u_j))$ for all u_j . Then the conditional distribution of each $(v^{(i)}(s, u_j) - v^{(i)}(t, u_j))$ is symmetric (but can be different for different i). Thus the i 'th profit of $T \in \mathcal{T}$ is $v^{(i)}(T) = \sum_{i \in [n-2]} (v^{(i)}(s, u_j) - v^{(i)}(t, u_j))(x_j - 1/2) + A^{(i)}$, 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 Cor. 7: For each fixing of $(v^{(i)}(s, u_j) + v^{(i)}(t, u_j))$, we get an instance of Cor. 7, and thus a lower bound of $(\frac{n-3}{2^{(d-1)}})^{d-1}$. Since this holds for each fixing of $(v^{(i)}(s, u_j) + v^{(i)}(t, u_j))$, we get that the same lower bound holds for the expectation without conditioning. ◀

4 0-1 Knapsack

We prove Theorem 3.

Proof. 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 $v^{(1)}, \dots, v^{(d)}$ 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 8 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 Improved lower bound in the semi-random model

We prove Theorem 4.

Proof. We only describe the differences with the argument in the proof of [5, Theorem 8]. As given by Theorem 3 (but scaling the profits by $1/\phi$, which does not change the set of Pareto-optima), we start with a distribution on knapsack instances with d profits and n_p objects (to be determined later) having unit weights and profits uniformly distributed in $[0, 1/\phi]$, and expected number of Pareto-optima at least $\Omega(n^{d-1.5})$. We use n_q (to be determined later) ‘‘cloning steps’’. Each step introduces d new objects while multiplying the number of Pareto-optima by at least $2^d/d$. As in [5], objects used by the splitting step can have profits that are larger than 1, therefore they are split into many objects with profits

distributed in $[0, 1]$, and a suitable choice of the set \mathcal{S} ensures that objects representing the split version of another behave as a group.

A simple modification of the argument leading to [5, Corollary 11], using our base case with n_p objects described in the previous paragraph instead of their base case with 1 object, implies that the expected number of Pareto-optima of the constructed instance is $\Omega(n_p^{d-1.5}(2^d/d)^{n_q})$.

Now we need to choose values of n_p and n_q to get a bound in terms of n . By [5, Lemma 11], the total number of objects is at most

$$n_p + dn_q + \frac{2d^2}{\phi - d} \left(\frac{2\phi}{\phi - d} \right)^{n_q} \quad (5)$$

We choose n_q so that the second term is no more than $n/4$ for n and ϕ sufficiently large. Such a choice of n_q is given by $n_q = \lceil \hat{n}_q \rceil$ with

$$\hat{n}_q = \frac{\log \phi}{\log \frac{2\phi}{\phi - d}}$$

when $4d^2 \leq n/4$ and $\phi \geq 2d$.

Clearly there can be no more than 2^n Pareto optima, and therefore there must be a point where increasing ϕ does not increase the lower bound. Say, for

$$\phi \leq \left(\frac{2\phi}{\phi - d} \right)^{n/2d} \quad (6)$$

we have that the second term in (5) is no more than $n/2$. Finally, choosing $n_p = \lfloor n/2 \rfloor$ ensures that (5) is at most n .

As explained in the first paragraph, the expected number of Pareto-optima of the whole construction is at least

$$\Omega \left(n_p^{d-1.5} \left(\frac{2^d}{d} \right)^{n_q} \right) \geq \Omega \left(n_p^{d-1.5} \left(\frac{2^d}{d} \right)^{\hat{n}_q} \right) \geq \Omega(n^{d-1.5} \phi^{(d-\log d)(1-\Theta(1/\phi))}).$$

When ϕ violates (6), we construct the same instance as above with maximum density equal to the unique $\hat{\phi}$ satisfying $\hat{\phi} = \left(\frac{2\hat{\phi}}{\hat{\phi} - d} \right)^{n/2d}$ ($\hat{\phi}$ is about $2^{n/2d}$). We get $\hat{n}_q = n/2d$ and, as before, the expected number of Pareto-optima is at least

$$\Omega \left(n_p^{d-1.5} \left(\frac{2^d}{d} \right)^{n_q} \right) \geq \Omega \left(n^{d-1.5} \left(\frac{2^d}{d} \right)^{n/2d} \right) \geq \Omega(2^{\Theta(n)})$$

◀

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 is of the form $\Omega(n^{d-1})$, ignoring the dependence on ϕ . The best upper bound we know, even for $\phi = 1$, is that of Moitra and O'Donnell [13] which is of the form $O(n^{2d-2})$, again 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 $\Theta_d(n^{d-1})$.

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 we need to use the technique of [5], which 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 lower bounding 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 [18]). 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 edge-isoperimetric inequality for the Boolean cube. Unfortunately, our numerical experiments show that this is not the case: Hamming balls are not the minimizers of the expected number of vertices of a random projection.

7 Acknowledgements

We thank Yusu Wang for helpful discussions. We are also very thankful to the Isaac Newton Institute for Mathematical Sciences, Cambridge, UK, where part of this work was carried out during the programme on Discrete Analysis.

References

- 1 René Beier, Heiko Röglin, and Berthold Vöcking. The smoothed number of pareto optimal solutions in bicriteria integer optimization. In *IPCO*, pages 53–67, 2007.
- 2 René Beier and Berthold Vöcking. Random knapsack in expected polynomial time. *J. Comput. Syst. Sci.*, 69(3):306–329, 2004.
- 3 René Beier and Berthold Vöcking. Typical properties of winners and losers in discrete optimization. *SIAM J. Comput.*, 35(4):855–881, 2006.
- 4 Jon Louis Bentley, H. T. 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.
- 5 Tobias Brunsch and Heiko Röglin. Lower bounds for the smoothed number of pareto optimal solutions. In *TAMC*, 2011.
- 6 Tobias Brunsch and Heiko Röglin. Improved smoothed analysis of multiobjective optimization. In *STOC*, 2012.
- 7 Tobias Brunsch and Heiko Röglin. Personal communication. June, 2011.

- 8 Christian Buchta. On the average number of maxima in a set of vectors. *Inf. Process. Lett.*, 33(2):63–65, 1989.
- 9 Luc Devroye. A note on finding convex hulls via maximal vectors. *Inf. Process. Lett.*, 11(1):53–56, 1980.
- 10 David Donoho and Jared Tanner. Counting the faces of randomly-projected hypercubes and orthants, with applications. *Discrete & Computational Geometry*, 43:522–541, 2010. 10.1007/s00454-009-9221-z.
- 11 J.E. Goodman and J. O’Rourke. *Handbook of discrete and computational geometry*. Discrete mathematics and its applications. Chapman & Hall/CRC, 2004.
- 12 Fabrizio Grandoni, R. Ravi, and Mohit Singh. Iterative rounding for multi-objective optimization problems. In *ESA*, pages 95–106, 2009.
- 13 Ankur Moitra and Ryan O’Donnell. Pareto optimal solutions for smoothed analysts. In *STOC*, 2011.
- 14 Matthias Müller-Hannemann and Karsten Weihe. Pareto shortest paths is often feasible in practice. In *Algorithm Engineering*, pages 185–198, 2001.
- 15 G. Nemhauser and Z. Ullmann. Discrete dynamic programming and capital allocation. *Management Science*, 15(9):494–505, 1969.
- 16 Franco P. Preparata and Michael Ian Shamos. *Computational geometry*. Texts and Monographs in Computer Science. Springer-Verlag, New York, 1985. An introduction.
- 17 Heiko Röglin and Shang-Hua Teng. Smoothed analysis of multiobjective optimization. In *FOCS*, pages 681–690, 2009.
- 18 Rolf Schneider. Recent results on random polytopes. *Boll. Un. Mat. Ital.*, Ser. (9), 1:17–39, 2008.
- 19 Rolf Schneider and Wolfgang Weil. *Stochastic and integral geometry*. Probability and its Applications (New York). Springer-Verlag, Berlin, 2008.
- 20 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.
- 21 J. G. Wendel. A problem in geometric probability. *Math. Scand.*, 11:109–111, 1962.