k-Pareto Optimality-Based Sorting with Maximization of Choice

Jean Ruppert
Mathematics and Computing S.à.r.l.
Luxembourg city, Luxembourg
jean.ruppert@mathcomp.lu

Marharyta Aleksandrova
University of Luxembourg
Esch-sur-Alzette, Luxembourg
marharyta.aleksandrova@uni.lu

Thomas Engel
University of Luxembourg
Esch-sur-Alzette, Luxembourg
thomas.engel@uni.lu

Abstract

Topological sorting[1] is an important technique in numerous practical applications, such as information retrieval, recommender systems, optimization, etc. In this paper, we introduce a problem of *generalized topological sorting with maximization of choice*, that is, of choosing a subset of items of a predefined size that contains the maximum number of equally preferable options (items) with respect to a dominance relation. We formulate this problem in a very abstract form and prove that sorting by *k-Pareto optimality* yields a valid solution. Next, we show that the proposed theory can be useful in practice. We apply it during the selection step of genetic optimization and demonstrate that the resulting algorithm outperforms existing state-of-the-art approaches such as *NSGA-II* and *NSGA-III*. We also demonstrate that the provided general formulation allows discovering interesting relationships and applying the developed theory to different applications.

1 INTRODUCTION

In the modern era of information overload, the task of choosing a subset of the most useful items is extremely important. Various tools were developed with the aim to assist a user with this task, for example, text search engines [Croft et al., 2010] and recommender systems

[Resnick and Varian, 1997]. In most of the cases, such systems suggest to the user a small set of elements[2]. Thereby, if the number of equally preferable options is large, a heuristic is used to discard a fraction of them. However, in some applications the user might be willing to analyze all equally preferable options with the aim to choose the best one. This can happen, for example, in the case of choosing a habitation.

A similar problem of *choosing a subset of most preferable elements* also arises as an important step when solving various practical tasks. A straightforward example would be the selection step in genetic optimization algorithms [Mitchell, 1998]. At this step, a subset of the current population is chosen to advance to the next generation. Having the chosen subset made up of elements with large fitness values guides the evolution process in the desired direction. At the same time, selecting a subset with the largest variety of genes ensures variability of characteristics and allows faster exploration of the search space.

These examples bring us to the problem of *generalized*[3] *topological sorting with choice maximisation* which we also refer to as *maximum choice problem*. This problem aims to choose a subset of a predefined maximum size, consisting of most preferable items and containing the largest number of equally preferred elements[4]. To the best of our knowledge, this problems has not yet been studied in the literature. In this text, we propose a theoretical solution to the maximum choice problem and demonstrate how both the problem and its solution can be applied in practice.

The contributions of this work are the following:

1. We formulate the maximum choice problem in a broad sense for arbitrary elements, preference re-

---

[1] *Topological sorting* [E.Knuth, 1997] here means the process of sorting a set of items with respect to a preference or dominance relation. We use the terms *preference* and *dominance* interchangeably.

[2] We use the terms *item* and *element* interchangeably.

[3] Later we show that the formulated problem is a generalization of topological sorting. In the case of a partial order relation, standard topological sorting is considered.

[4] If items are equally preferable, then we say that they offer choice for the user or the system.

lations R, and measures μ indicating the set sizes, see Section 2.

2. We propose a solution based on the concept of k-Pareto optimality, whose definition relies on the relation R, see Section 3.

3. We further investigate the proposed solution from the theoretical point of view and discover interesting characteristics, such as the relationship between k-Pareto optimal elements and the arc of hyperbola, see Section 4.

4. Finally, we demonstrate the applicability of our approach to real-world problems by considering genetic optimization, see Section 5.

2 SORTING WITH CHOICE MAXIMIZATION

To formally define the problem of *maximization of choice*, we introduce several definitions in Sections 2.1 and 2.2. The resulting formalization is abstract and quite general. However, this generality allows discovering novel connections and applying the developed theory to numerous practical problems. We also illustrate the defined concepts with an example in Section 2.3.

2.1 Definitions

We consider a set X with a binary relation R. Intuitively xRy means that x is preferable to y. The case xRy and not yRx means that x is *strictly preferable* to y. This situation is denoted by xR^*y. We also consider a positive and σ-finite measure μ [Halmos, 2013] defined on X. Thus, we have a measure space (X, Σ, μ), where Σ is a set of subsets of X, and μ intuitively indicates the size of these subsets. To ensure measurability, throughout this text the characteristic function 1_R of the relation R is assumed to be sufficiently regular[5]. The measure μ can be defined in different ways. Important examples are the counting measure and probability measure P. Depending on the definition of μ, it can indicate the following characteristics of the elements in X: *how many?*, *how likely?*, *how important?*, or *what volume?*

To illustrate these definitions, we consider the following example. Let X be a set of possible habitations of which the user has to choose the best according to his preferences encoded by the relation R. In such a situation, the relation R can be multidimensional. Let us assume, for simplicity, that an optimal habitation for the user is close to a given location, for example, his workplace (relation R_l), is situated in a district

with a smaller population size (relation R_p), and is close to a river (relation R_r). Thus, the user's preferences can be represented by the preorder relation $R = R_l \& R_p \& R_r$. In our example, all available habitations from X can be mapped onto points in a 3-dimensional space of *Proximity to the location* $\times$ *Population* $\times$ *River*. The fact that R is a preorder relation means that some elements of X can be comparable, while others not. For example, the habitation x with coordinates $(50, 100, True)$ is strictly preferable to y with coordinates $(60, 100, True)$, that is xR^*y. At the same time, the habitation z with coordinates $(40, 100, False)$ is incomparable with x. Indeed, z is better with respect to R_l, it is situated closer to the required location, but x is better with respect to R_r, as the latter is situated near a river.

Having the task to find a subset of X that is 'best' according to R, a rational solution can be formulated with the following recursive expression: *if an element x is selected, then all elements that are strictly preferable to x should be also selected*. In our example, this translates into the task of finding a subset of habitations S_R that might be suitable for the user. Naturally, if $y \in S_R$, then $x \in S_R$ as the latter corresponds better to the preferences of the user defined by the relation R. We formalize this rationality condition by defining *selections* as follows.

Definition 1. *A selection S is a subset of X such that $x \in S$ and yR^*x implies $y \in S$. The set of all selections in Σ is denoted by $\mathcal{S}$*

If R is a partial order relation, then the selections are the down-sets [Davey and Priestley, 2002]. A selection of size n is obtained by taking the n best items according to a linear extension of R. The number of linear extensions of a relationship is often very large.

2.2 The Maximum Choice Problem

As discussed in Section 1, in practical applications when selecting a subset of X one might want not only to respect the above rationality constraint, but also to maximize the number of incomparable pairs. The latter condition is equivalent to the maximization of the diversity of the selected subset, or the maximization of the provided choice. In our example with habitations, if both x and z are presented to the user, then he can choose an appropriate habitation by himself[6].

In terms of our notations, this will be translated into the condition of selecting as many pairs x, y such that neither x is strictly preferable to y ($\neg xR^*y$) nor y is strictly preferable to x ($\neg yR^*x$). This means that

---

[5]1_R is equal to 1 if xRy and is equal to 0 otherwise.

[6]We consider the case when addition preferences cannot be encoded and the user has to make the final choice.

there is freedom of choice between x and y ($xRy = yRx$). This motivates the following quantitative definition of choice for measurable subsets of X.

Definition 2. Choice *offered by a set A is the number* $\text{cho}(A) = (\mu \times \mu)(\{(x, y) \in A^2 | xRy = yRx\})$, *where* $(\mu \times \mu)$ *is the product measure [Halmos, 2013].*

The choice offered by a measurable set A essentially measures how many pairs of items offering choice can be extracted from A. Additionally, if one wants to restrict the size of the selected subset, in our example, to present to the user a small set of suitable habitations with $\mu(S_R) \leq m$, then this leads us to the definition of the maximum choice problem:

Maximum Choice Problem. *For a given m find all selections T such that* $\text{cho}(T) = \max_{S \in \mathcal{S}, \mu(S) \leq m} \text{cho}(S)$. *Any such T is said to* offer maximum choice for m.

The main ingredient of our solution to the maximum choice problem is the following concept.

Definition 3. k-Pareto optimality[7] *of an element $x \in X$ is the measure of the subset of X containing all elements strictly preferable to x:* $\text{po}(x) = \mu(\{y | yR^*x\})$.

If μ is the probability measure, the k-Pareto optimality of an element x, $\text{po}(x)$, is the likelihood an element drawn at random from X is strictly preferable to x. In Section 3, we demonstrate that the following sets yield a solutions for the *maximum choice problem.*

Definition 4. *The* at least k-Pareto optimal elements T_k *form the measurable set defined as follows:* $T_k = \{x \in X | \text{po}(x) \leq k\}$.

If R^* is transitive, then for any k, T_k is a selection.

2.3 Example

In this subsection, we discuss an illustrative example to demonstrate the concepts defined in Sections 2.1 and 2.2. Let us consider a finite subset X of $\mathbb{R}^2$, the counting measure μ, and the relation $R_\llcorner$ defined as follows: $(x_1, x_2)R_\llcorner(y_1, y_2)$ iff $x_1 \leq y_1$ and $x_2 \leq y_2$, see Fig. 1[8]. In economics, x is Pareto optimal if there is no y in X such that $yR^*_\llcorner x$. In our language, this means that $\text{po}(x) = 0$. Thus, k-Pareto optimality indicates how much an element is away from being Pareto optimal.

Let X be comprised of six points presented in Fig. 2. As we are considering the counting measure, $\mu(X) = 6$. Points A, B, and C are not dominated by any other

---

[7]The value of k is directly used only in Def. 4. However, we keep it for consistency of the developed theory.

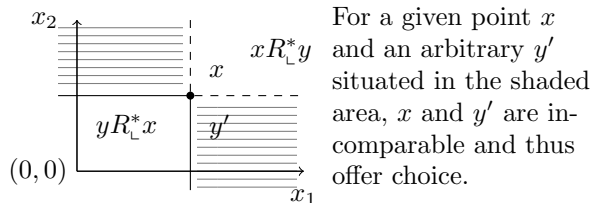[8]See Appendix A for examples with other measures and relations.

For a given point x and an arbitrary y' situated in the shaded area, x and y' are incomparable and thus offer choice.

Figure 1: Illustration of the relation $R_\llcorner$.

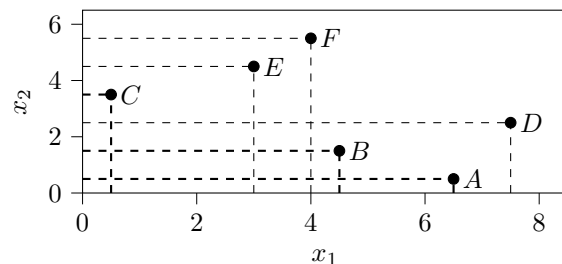
Figure 2: Computation of Pareto-optimality. Example of a finite subset of $\mathbb{R}^2$, the counting measure, and $R_\llcorner$.

point. This means that $\text{po}(A) = \mu(\{y | yR^*A\}) = 0$ and $\text{po}(B) = \text{po}(C) = \text{po}(A) = 0$. Point E is dominated by a single point C, that is $\text{po}(E) = \mu(\{C\}) = 1$. Finally, points F and D are dominated by two other points each, resulting in $\text{po}(F) = \mu(\{C, E\}) = 2$ and $\text{po}(D) = \mu(\{A, B\}) = 2$.

Sorting the set X by k-Pareto optimality of its elements will produce the following result: $(\{A, B, C\}, \{E\}, \{D, F\})$. This sorting is different from sorting by Pareto fronts. The latter approach is widely used in practice and is the basis of all Pareto dominance-based genetic optimization algorithms [Li et al., 2015]. Sorting by Pareto fronts is done in the following way. First, the first Pareto front, which is the set of non-dominated points, is identified. Next, the points from this front are removed from the consideration and the process is repeated until all points are assigned to a front. Sorting the points from Fig. 2 by Pareto fronts will produce the following result: $(\{A, B, C\}, \{D, E\}, \{F\})$. Note, that the point D moved from the 3d equivalence class when sorting by k-Pareto optimality to the 2d when sorting by Pareto fronts. Let us consider the two selections of size 4: $S_E = \{A, B, C, E\}$ and $S_D = \{A, B, C, D\}$. Sorting by Pareto fronts does not distinguish between these two selections as both E and D belong to the same equivalence class. At the same time, sorting by k-Pareto optimality has a larger preference towards S_E. Also, the latter selection contains more incomparable pairs of elements and thus offers more choice. Indeed, D is incomparable with only one point C, but E is

incomparable with two points: A and B[9].

For the counting measure and a discrete space, sorting by k-Pareto optimality is identical to sorting by *dominance rank* [Liefooghe et al., 2009]. This sorting procedure was studied in genetic optimization [Reyes-Fernández-de Bulnes et al., 2019]. However, to the best of our knowledge, it was never considered beyond this setting. Also, its relation to the *maximum choice problem* is novel.

3   MAXIMUM CHOICE THEOREM

The main result of this text states that at least k-Pareto optimal elements (T_k) are the largest measurable selections offering maximum choice for their respective measures. In Section 3.1, we state and prove the maximum choice theorem, and in Section 3.2 we discuss further solutions.

3.1   Theorem and Proof

Theorem 1 (Maximum Choice Theorem). *A set of at least k-Pareto optimal elements T_k offers maximum choice for $\mu(T_k)$ if it is a selection and if $\mu(T_k) < +\infty$. Moreover, T_k is the largest selection offering maximum choice for $\mu(T_k)$ in the sense that it contains any other selection offering maximum choice for $\mu(T_k)$.*

The second part of the above theorem precisely means that if a selection A offers maximum choice for $\mu(T_k) < +\infty$, then $A \subseteq T_k$ almost-everywhere. The above theorem does not state that the at least k-Pareto optimal elements are the only largest selections offering maximum choice. Later, we give an example of a fundamentally different selection offering maximum choice for a value of m where there is no such k that $\mu(T_k) = m$.

We prove the above theorem in several steps. First, we show that for selections the computation of choice can be simplified. It only requires to compute a simple integral instead of a double integral.

Integral Formula. *If S is a measurable selection and $\mu(S) < +\infty$, then*

$$\mathrm{cho}(S) = \int_S (\mu(S) - 2\,\mathrm{po}(x))d\mu(x). \tag{1}$$

Proof. The fact that set S is a selection means that $\forall y \in S : \{x \in S | xR^*y\} = \{x \in X | xR^*y\}$. That is, any element x from X strictly preferable to any element y in S, also belong to S ($x \in S$). Using the definition of

choice from Def. 2 and $\mu(S) < +\infty$, we obtain

$$\mathrm{cho}(S) = \mu(S)^2 - 2(\mu \times \mu)(\{(x,y) \in S^2 | xR^*y\})$$
$$= \mu(S)^2 - 2(\mu \times \mu)(\{(x,y) \in S \times X | yR^*x\}).$$

Fubini's theorem [Halmos, 2013] indicates that

$$(\mu \times \mu)(\{(x,y) \in S \times X | yR^*x\} =$$
$$= \int_S \left(\int_X 1_{R^*} d(\mu(y)) \right) d\mu(x) = \int_S \mathrm{po}(x)d\mu(x),$$

where 1_{R^*} is the characteristic function of R^*[10].

Finally, the integral formula results from the fact that $\mu(S)^2 - 2\int_S \mathrm{po}(x)d\mu(x) = \int_S (\mu(S) - 2\,\mathrm{po}(x))d\mu(x)$.   □

Let's now consider the function c defined on Σ for any A of finite measure by

$$c(A) = \int_A (\mu(A) - 2\,\mathrm{po}(x))d\mu(x).$$

The integral formula defined in Eq. (1) means that for any selection S, we have $c(S) = \mathrm{cho}(S)$. The second step of our proof of Theorem 1 is to show that T_k is the largest measurable set that maximizes c for its respective measure. We prove the following lemma.

Lemma 1. *For any k such as $\mu(T_k) < +\infty$ we have*

$$c(T_k) = \max_{A \in \Sigma, \mu(A) \leq \mu(T_k)} c(A).$$

Moreover, if $\mu(A) \leq \mu(T_k)$ and $c(A) = c(T_k)$, then $A \subseteq T_k$ almost-everywhere.

The context of this lemma is very similar to the knapsack problem [Martello, 1990]. In this problem, one needs to find a subset A of a finite set of items $\{x_1, ..., x_n\}$ maximizing the total value $\Sigma_{x_i \in A} v(x_i)$ under the constraint that the total weight $\Sigma_{x_i \in A} w(x_i)$ of A does not exceed a predefined maximum weight w^*. In Lemma 1, the total value is $c(A)$, the ratio of an element's value to its weight becomes $\mu(A) - 2\,\mathrm{po}(x)$, and the weight constraint is expressed as $\mu(A) \leq \mu(T_k)$. The solutions given by the lemma correspond to those yielded for the knapsack problem by George Dantzig's greedy approximation algorithm [Dantzig, 1957]. This algorithm consists of ordering elements by decreasing value-to-weight ratio and then taking the N first elements. N is chosen in such a way, that taking one more element would cause excessive weight. The process of proving Lemma 1 is similar to proving that George Dantzig's solutions are optimal for their respective weights.

---

[9]The difference between sorting by k-Pareto optimality and Pareto fronts is further discussed in Appendix B.

[10]1_{R^*} is defined on X^2 in a similar way to 1_R: it is equal to 1 if xR^*y and is equal to 0 otherwise.

Proof of Lemma 1. Let's consider T_k such that $\mu(T_k) < +\infty$. To prove $c(T_k)$ is the maximum, we need to show for any $A \in \Sigma$, that $c(A) \leq c(T_k)$ if $\mu(A) \leq \mu(T_k)$. As $A = (A \cap T_k) \cup (A \setminus T_k)$ and $T_k = (A \cap T_k) \cup (T_k \setminus A)$, requiring $c(A) \leq c(T_k)$ is equivalent to requiring

$$\int\limits_{A \setminus T_k} (\mu(A) - 2\operatorname{po}(x))d\mu(x) \leq \int\limits_{T_k \setminus A} (\mu(T_k) - 2\operatorname{po}(x))d\mu(x). \tag{2}$$

By definition of T_k, we have that $\operatorname{po}(x) > k$ for $x \in A \setminus T_k$, while $\operatorname{po}(x) \leq k$ for $x \in T_k \setminus A$. Therefore,

$$\int\limits_{A \setminus T_k} (\mu(A) - 2\operatorname{po}(x))d\mu(x) \leq \mu(A \setminus T_k)(\mu(A) - 2k), \tag{3}$$

and

$$\int\limits_{T_k \setminus A} (\mu(T_k) - 2\operatorname{po}(x))d\mu(x) \geq \mu(T_k \setminus A)(\mu(T_k) - 2k). \tag{4}$$

The constraint $\mu(A) \leq \mu(T_k)$ means that $\mu(A \setminus T_k) \leq \mu(T_k \setminus A)$. Thus, we have

$$\mu(A \setminus T_k)(\mu(A) - 2k) \leq \mu(T_k \setminus A)(\mu(T_k) - 2k). \tag{5}$$

Finally, combining Eq. (5) with Eq. (3) and Eq. (4) guarantees the inequality in Eq. (2) under the constraint $\mu(A) \leq \mu(T_k)$.

Let us now consider $A \in \Sigma$ such that $\mu(A) \leq \mu(T_k)$ and $c(A) = c(T_k)$. We now proceed to show that $A \subseteq T_k$ almost everywhere. If $c(A) = c(T_k)$, the inequality in Eq. (2) must be an equality. Under the assumption $\mu(A) \leq \mu(T_k)$, we again have the inequalities in Eq. (3), Eq. (4) and Eq. (5), which must be equalities if Eq. (2) is an equality. However, because $\operatorname{po}(x) > k$ for $x \in A \setminus T_k$, the inequality in Eq. (3) can only become an equality if $\mu(A \setminus T_k) = 0$. □

Now everything is in place to prove the theorem.

Proof of Theorem 1. Let us first prove that the at least k-Pareto optimal elements T_k offer maximum choice. Lemma 1 says that on Σ the set T_k maximises c for its respective measure. We assumed T_k is a selection. At the same time, for any selection $c = \operatorname{cho}$. It means that T_k offers maximum choice for its respective measure. Moreover, from Lemma 1 and from $\mathcal{S} \subseteq \Sigma$ directly results that if a selection A offers maximum choice for $\mu(T_k)$, then $A \subseteq T_k$ almost everywhere. □

Uniqueness and transitivity. The theorem does not guarantee *uniqueness*. For example, in the case of the relation $\leq$ on $\mathbb{R}$ and the Lebesgue measure, selections are all left-unbounded intervals and the choice of any selection is 0. However, if for any selection

$$A \subset T_k \ \& \ \mu(A) < \mu(T_k) \implies \operatorname{cho}(A) < \operatorname{cho}(T_k), \tag{6}$$

then T_k is a unique maximum. This is a direct consequence from the fact that T_k contains any other selection offering maximum choice for $\mu(T_k) < +\infty$.

The theorem requires the set T_k only to be a selection. If R is *transitive*, then T_k is always a selection. The set T_k already fulfills the requirement of being selection if R^* satisfies the weakened transitivity condition $xR^*y \implies \operatorname{po}(x) \leq \operatorname{po}(y)$. This is the case for Lebesgue area measure and the non-transitive relation $R_{\llcorner'}$ defined on the unit square by $(x_1, x_2)R_{\llcorner'}(y_1, y_2)$ iff $\frac{y_1}{2} \leq x_1 \leq y_1$ and $\frac{y_2}{2} \leq x_2 \leq y_2$.

3.2 Further Solutions

Further similar solutions. It is possible to prove that Theorem 1 also holds for T_k^* defined with a strict inequality $(<)$ as follows, see Def. 4 for comparison.

$$T_k^* = \{x \in X \,|\, \operatorname{po}(x) < k\}.$$

In this case, the proof of the fact that T_k^* is the largest in Lemma 1 requires analysis of inequality (4) instead of inequality (3). However, the proof of the fact that any selection T such that $T_k^* \subseteq T \subseteq T_k$ offers maximum choice for $\mu(T)$ becomes a bit more technical. Moreover, it is possible to prove that T is the largest selection of this kind. Precisely, for any other selection S offering maximum choice for $\mu(T)$, one must have that $S \subseteq T'$ for some T' such that $\mu(T') = \mu(T)$ and $T_k^* \subseteq T' \subseteq T_k$.

Completeness. In the case condition from Eq. (6) holds, and if for any selection S there is a k and there are selections T such that $T_k^* \subseteq T \subseteq T_k$ and $\mu(T) = \mu(S)$, then those selections T are the only selections offering maximum choice. Therefore, we have a complete list of selections offering maximum choice. This is the case for the typical example of the relation $R_\llcorner$ defined in Section 2.3 and the Lebesgue area measure defined on the unit square $[0,1]^2$. This case is further studied Section 4.

The above ideas also apply to any finite set on which we have a partial order relation and discrete non-zero weights. Topological sorting by increasing values of po and then taking the first N elements yields a set offering maximum choice. If there are several ties, that is if $T_k \setminus T_k^*$ contains several elements, then finding all sets T of the allowed maximum measure such as $T_k^* \subset T \subset T_k$ means solving the generally NP-complete subset-sum problem. If, however, the weights are constant, any
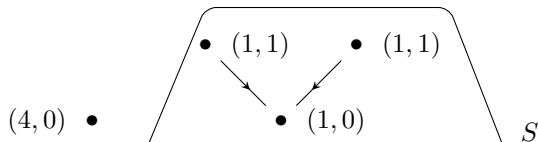
Figure 3: S is a selection that offers maximum choice but cannot be constructed from T_k. Point are encoded with their measure and k-Pareto optimality: (μ, po).

(a) x and y offer choice.   (b) y is preferable to x.

Figure 4: Characterization of hyperbola.

set offering maximum choice can be obtained using the above sorting procedure; with ties being ordered in an arbitrary way.

Existence of solutions of a different nature. For discrete measures with non-constant weights the above construction might not yield all selections offering maximum choice. A relative example is given in Fig. 3, where partial order relation is represented by its Hasse diagram [Davey and Priestley, 2002]. The set S is not of the form $T_k^* \cup A$ with $A \subseteq T_k \setminus T_k^*$. Nevertheless, it offers maximum choice for $m = 3$.
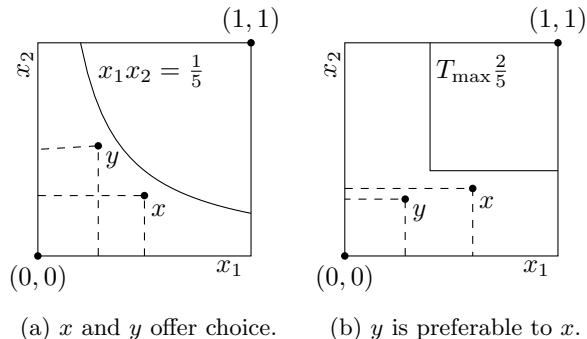
4 THEORETICAL EXPLORATIONS

The developed theory is very general as there are no restrictions on X and μ, and R only needs to fulfill a weakened transitivity condition. This makes it applicable to both discrete and continuous examples. In this section, we further investigate the theoretical properties of sorting by k-Pareto optimality. We focus on the relation $R_{\llcorner}$ defined on $\mathbb{R}^2$. This is done to gain an intuitive understanding in this familiar setting from the point of view of geometry and probability theory. The theoretical analysis presented in this section will be extended in future work to less familiar settings, for example, the subgraph relation defined on a set of graphs.

This section is structured as follows. First, we demonstrate the link between k-Pareto-optimality and the concept of diversity in Section 4.1. Next, in Section 4.2, we analyze the continuous case and show how the developed theory can provide a surprising characterization of the arc of a hyperbola. In Section 4.3, we continue with the case of continuous random vectors and compare our sorting method with other well-known ways of topologically sorting $\mathbb{R}^2$. Finally, in Section 4.4, we discuss how the analysis of continuous probabilistic examples can be used in practice.

4.1 Diversity

Choice turned out to be a natural concept when developing the above theory and proving Theorem 1. In practical applications, however, it can be more insight-

ful to consider the following concept.

Definition 5. *For any measurable set A, the* diversity *of A is the ratio*

$$\text{div}(A) = \frac{\text{cho}(A)}{\mu(A)^2}.$$

Thus, $\text{div}(A)$ is the likelihood there is choice between two elements chosen at random from A.

In Theorem 1, we cannot simply replace choice with diversity. However, by considering only selections of a fixed measure, we obtain the following straightforward corollary of Theorem 1.

Corollary 1. *For any set of at least k-Pareto optimal elements T_k that is a selection with $\mu(T_k) < +\infty$, we have*

$$\text{div}(T_k) = \max_{S \in \mathcal{S}, \mu(S) = \mu(T_k)} \text{div}(S).$$

Moreover, T_k is the unique such maximum. Precisely, if S is a selection such that $\mu(S) = \mu(T_k)$, and $\text{div}(S) = \text{div}(T_k)$, then $S = T_k$ almost everywhere.

4.2 Characterisation of the Hyperbola

Let us consider X to be the unit square $[0, 1] \times [0, 1]$ on which we have the Lebesgue area measure and the relation $R_{\llcorner}$. For any point $x = (x_1, x_2)$ in X, the k-Pareto optimality of x is the area of the rectangle $((0, 0), (x_1, 0), x, (0, x_2))$. Thus, $\text{po}(x) = x_1 x_2$ and the at least k-Pareto optimal elements are situated below the hyperbola $x_1 x_2 = k$. Selections are sets situated below any decreasing curve, for example, the hyperbola $x_1 x_2 = \frac{1}{5}$ and the curve defined by $\max(x_1, x_2) = \frac{2}{5}$, see Fig. 4. Having choice between x with coordinates (x_1, x_2) and y with coordinates (y_1, y_2) means the rectangles $((0, 0), (x_1, 0), x, (0, x_2))$ and $((0, 0), (y_1, 0), y, (0, y_2))$ are not nested, as depicted in Fig. 4a. Combining this fact with Corollary 1 yields the following surprising characterization of the hyperbola.

Characterization of Hyperbolas. *Out of all descending functions f from $[0,1]$ to $[0,1]$ delimiting an area $\int_0^1 f(x)dx = c$, the arc of hyperbola is the one offering the highest likelihood the rectangles $((0,0),(x_1,0),x,(0,x_2))$ and $((0,0),(y_1,0),y,(0,y_2))$ are not nested for two points x and y being drawn independently and at random from the delimited area.*

4.3 Choice and Aggregation of Ranks

Let's consider a continuous random vector (X_1, X_2). According to the standard notation, values taken by (X_1, X_2) will be written in lower case, for example (x_1, x_1) or (y_1, y_2). Let us again consider $R_\llcorner$ giving precedence to smaller values of X_1 and X_2.

Continuity means $P(\mathrm{po}((x_1, x_2)) = \mathrm{po}((y_1, y_2))) = 0$. Thus, the function po can be considered to be a linear extension of $R_\llcorner$. On the other hand, in the case of independence between X_1 and X_2 the diversity of the whole probability space is $\frac{1}{2}$. This means, that only in half of the cases $R_\llcorner$ can directly tell which of the two randomly chosen elements is preferable.

In Appendix B.2, we show that k-Pareto optimality, choice, diversity and selections offering maximum choice only depend on the copula of (X_1, X_2), which describes the dependence structure between X_1 and X_2, see Proposition 2. Moreover, $\mathrm{po}((x_1, x_2))$ is the joint cumulative probability distribution function of the random vector (X_1, X_2). In probability theory the uniform distribution on the unit square from Section 4.2 is the copula of any two continuous independent random variables. Then, x_1 and x_2 represent the ranks of two independent characteristics, and sorting by k-Pareto optimality means sorting by increasing value of the product $x_1 x_2$.

Assuming X_1 and X_2 are independent, and using the change of variable formula from Proposition 1 (Appendix B.1), we show in Proposition 3 (Appendix B.2) that for at-least k-Pareto optimal elements T_k

$$P(T_k) = k - k \ln(k),$$

$$\mathrm{cho}(T_k) = (k - k \ln(k))^2 - k^2 \left(\frac{1}{2} - \ln k \right).$$

This implies that $\lim_{k \to 0} \mathrm{div}(T_k) = 1$. It means that diversity slowly tends to the maximum possible value of 1 as k tends to zero. This fact becomes even more surprising when comparing to other methods of aggregating ranks that yield linear extensions of $R_\llcorner$, for example, minimum rank, maximum rank and mean rank. Let us look at the corresponding selections

$$T_{\min a} = \{(x_1, x_2) \in [0,1]^2 \,|\, \min(x_1, x_2) \le a\},$$

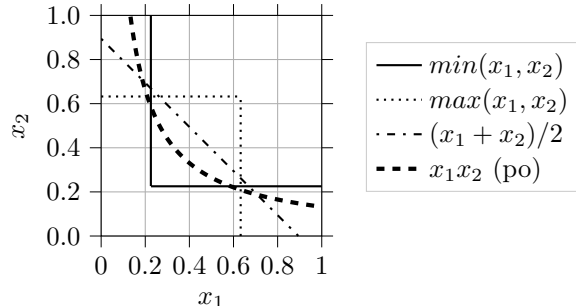$$T_{\max a} = \{(x_1, x_2) \in [0,1]^2 \,|\, \max(x_1, x_2) \le a\}.$$

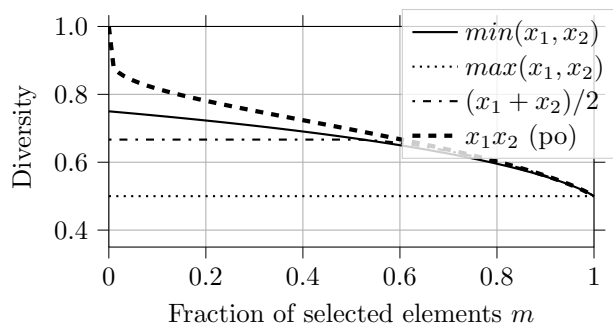Figure 5: Selections of the best 40% according to different sorting criteria.

Figure 6: Fraction of the best elements and diversity.

Here, $\lim_{a \to 0} \mathrm{div}(T_{\min a}) = \frac{3}{4}$, and $\mathrm{div}(T_{\max a}) = \frac{1}{2}$. To further illustrate this observation, let us consider selections of a fixed measure m, which represent the best $m * 100\%$ of elements, defined with different sorting criteria: minimum $min(x_1, x_2)$, maximum $max(x_1, x_2)$, average $\frac{x_1 + x_2}{2}$, and Pareto optimality $x_1 x_2$[11]. As we can see in Fig. 5, $min(x_1, x_2)$ delimits selections containing too many large values, that is, extremes are overvalued. On the other hand, all other sorting criteria except po undervalue extremes and include too many elements situated around the diagonal $x_1 = x_2$.

Finally, in Fig. 6 we demonstrate how diversity of the selections defined above depends on the fraction of selected elements m[12]. We can see that diversity is the largest when sorting by k-Pareto optimality. This is a direct consequence of Corollary 1.

4.4 Limiting Behaviour and Probabilistic Framework

The continuous probabilistic examples considered in Sections 4.2 and 4.3 might seem unpractical and interesting only from the theoretical point of view. How-

---

[11]Here, po is functionally related to geometric mean.
[12]In the considered case, for any selection S, $\mathrm{cho}(S) = dx_1 dx_2 (S)^2 - \int_S \mathrm{po}(x) dx_1 dx_2$.

ever, such a case may arise if we have a very large population that can be approximated by a continuous distribution. Then, in real-world applications, one may use the best adapted parametric and non-parametric statistical methods to estimate *k*-Pareto optimality. This is the idea behind the approximation algorithm presented in Section 5.1 and the computationally efficient approximation formula in Eq. (7).

Moreover, measure theoretic concepts can be used to formulate and prove convergence of *k*-Pareto optimality, choice and selections as the population becomes larger and its distribution converges to a continuous distribution. On the other hand, topological sorting by Pareto fronts demonstrates unstable behaviour, as shown in Appendix B.3: it is more sensitive to the precise way points are placed on the plane. In the last example of this section, we also show how the selections obtained via sorting by Pareto fronts can diverge even when the selections obtained via sorting by *k*-Pareto optimality converge to the arcs of hyperbola obtained in Section 4.2.

5 PRACTICAL EXPLORATIONS

In this section, we discuss computational complexity of sorting by *k*-Pareto optimality, see Section 5.1, and demonstrate how the proposed theory can be used to improve the performance of genetic optimization algorithms, see Section 5.2. Further, in Appendix C, we discuss other potential applications, such as statistical tests (Appendix C.2), recommender systems (Appendix C.3), constrained genetic optimization (Appendix C.4), exploratory database queries (Appendix C.5), and scheduling (Appendix C.6).

5.1 Computation Complexity

A common solution for ranking n elements of a set X according to a partial order relation R is to rank the elements according to their average ranking with respect to all linear extensions of R. However, the total number of linear extensions exponentially increases with n, and the resulting algorithms are complex and slow [de Loof, 2010, p. 48]. For example, random sampling of linear extensions has an expected running time of $O(n^3 \log n)$ [Huber, 2006]. Below we show that sorting by *k*-Pareto optimality offers an efficient alternative.

The basic algorithm for the *k*-Pareto optimality based sorting is straightforward. In the case of an arbitrary relation R, po(x) is computed by summing up the measures of the items that are strictly preferable to x. This requires one pass through the whole set X for every element $x \in X$ with computation complexity $O(n^2)$.

The complexity of sorting X by increasing values of po is $O(n \log n)$. Therefore, the total complexity is $O(n^2)$.

The case of composite relations defined on the probability space allows constructing even faster sorting procedures. We illustrate this idea for $R = R_l \& R_p \& R_r$ from our housings example. We define the component relations as follows: for $i \in \{l, r, p\}$, $a R_i b$ iff $X_i(a) \leq X_i(b)$, where the real valued random variable X_l represents proximity to the location, X_p represents population size, and $X_r(x)$ is 0 when x is close to a river and 1 otherwise. For independent X_i, we have:

$$
\begin{aligned}
\text{po}(x) &= P(\{y | y R^* x\}), \\
&= P(\{y | y R x\}) - P(\{y | y R x \text{ and } x R y\}), \\
&= \prod_{i \in \{1,...,m\}} P(X_i \leq x_i) - \prod_{i \in \{1,...,m\}} P(X_i = x_i). \quad (7)
\end{aligned}
$$

The cumulative probability distributions $F_i(x) = P(X_i \leq x)$ can be approximated by the respective empirical cumulative probability distributions $\hat{F}_i(x)$. The computation complexity of estimating $\hat{F}_i$ is $O(n \log n)$. This needs to be done for every component relation R_i, resulting in the total complexity of $O(n \log n)$.

5.2 Application to Genetic Optimization

In Section 1, we hypothesized that sorting with choice maximization can be beneficial for genetic optimization. Indeed, this strategy results in the maximization of the population diversity and allows exploring the search space more efficiently. Additionally, Pareto dominance-based many-objective[13] genetic optimization algorithms are known to suffer from the lack of selection pressure [Palakonda et al., 2018]. When the number of objectives increases, the number of incomparable solutions grows exponentially. However, as shown in Section 4.3, sorting random independent vectors by their Pareto optimality can be considered as a linear extension of the defined preference relation. The fact that $P(\text{po}(x) = \text{po}(y)) = 0$ means that such sorting rarely produces ties and for any two solutions either x is preferable to y or vice versa. In the rest of this subsection, we demonstrate that the proposed approach indeed improves the performance of genetic algorithms in the case of independent objectives.

To evaluate the proposed sorting procedure, we use it in *NSGA-II* instead of Pareto dominance-based sorting. We experiment with two measures μ: counting and probability measures. This gives us two versions of genetic algorithms referred to as *PO-count* and *PO-prob* respectively. In *PO-count*, the solutions are sorted according to the number of other dominating solutions. As discussed in Section 2.3, in this case,

---

[13]Concerns problems with 4 and more objectives.

sorting by k-Pareto optimality is identical to sorting by *dominance rank*. Thereby, these results are not novel, but we present them for comparison purposes. In *PO-prob*, the number of dominating solutions is replaced with the probability of being dominated. We consider a given generation as a small sample of a much larger population having the same marginal probability distributions. We assume there is no a-priori knowledge on how objectives are correlated and, thus, assume them to be independent. The probability of being dominated in this much larger population is estimated as explained in Section 5.1. These algorithms, *PO-count* and *PO-prob*, are compared with implementations of the state-of-the-art algorithms *NSGA-II* and *NSGA-III* [Deb and Jain, 2013] from *deap* python library[14]. The presented experiments as well as various plots from this paper can be reproduced using the code from the related GitHub repository[15].

For the experimental evaluation, we use the 0/1 knapsack problem with independent objectives as defined in [Zitzler and Thiele, 1999]. The number of knapsacks (objectives) is varied within the following set $n_k \in \{2-8, 10, 15, 25\}$ and the number of items is set to 250. We adopt random selection with replacement and uniform crossover with mutation probability 0.01. We set the population size to 250 and the number of generations to 500. All results are the average among 30 independent runs.

Below we analyze the performance of different algorithms in terms of the classical *hypervolume* metric [Shang et al., 2020] with the origin of coordinates as a reference point. In our setup, this metric is to be maximized. We choose *NSGA-II* as the baseline, and present the relative changes in the **hypervolume** indicator for the rest of the algorithms in Fig. 7 (increase: positive number, decrease: negative number). We notice that despite having been developed for the many-objective optimization, *NSGA-III* almost always results in lower values of hypervolume, even for a large number of knapsacks. This confirms a similar observation from [Ishibuchi et al., 2016], and supports our choice of *NSGA-II* as a baseline for implementation and comparison instead of *NSGA-III*. Further, we see that the value of relative increase for *PO-count* is always very close to 0. It means that *PO-count* yields a population covering the same hypervolume as *NSGA-II*. Contrarily, *PO-prob* improves the hypervolume, as compared to *NSGA-II*. This difference is visible for small n_k (+4% for $n_k = 2$) and is especially prominent for large n_k (+60% for $n_k = 25$). For n_k between 5 and 7, *PO-prob* results in lower values of hypervolume than *NSGA-II*. However, the relative decrease in
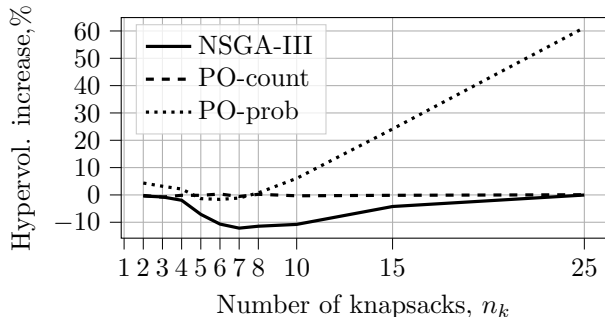

Figure 7: Increase in hypervolume compared to *NSGA-II*.

these cases does not exceed -1.63%. Also, within this range, *PO-count* performs slightly better than other algorithms. These results demonstrate that the proposed approach improves the performance of genetic algorithms, especially in the case of many-objective optimization. Additional experimental results presented in Appendix C.1 also support this statement. Our results also suggest that the choice of the measure μ has a large impact on the performance. The latter relationship will be studied in future work.

6 CONCLUSION

In this paper, we formulate the problem of *generalized topological sorting with choice maximization*, which, to the best of our knowledge, was not considered in the literature before. We also prove that the *at least k-Pareto optimal sets* provide unique solutions. Further theoretical analysis of this problem leads us to an interesting relationship between the diversity of random points and the arc of hyperbola. Additionally, we propose a computationally efficient algorithm for calculation of k-Pareto optimality for probability measures. Finally, we demonstrate a successful application of the developed theory. We show that sorting by k-Pareto optimality can drastically improve the performance of many-objective genetic optimization algorithms. In our experiments, the proposed solution based on the probability measure allows increasing the value of hypervolume by up to 60% for 25 objectives. This result can be considered as a potential solution to the problem of searchability deterioration in Pareto-dominance optimization.

We also believe that the proposed general framework can be used in different applications. In future work, we plan to study the applicability of k-Pareto optimality for constrained optimization, scheduling problems, recommender systems, and the development of statistical indicators. Maximization of choice might be also useful when studying causality and fairness.

References

[Billingsley, 1999] Billingsley, P. (1999). *Convergence of Probability Measures*. Springer, 2 edition.

[Croft et al., 2010] Croft, W. B., Metzler, D., and Strohman, T. (2010). *Search engines: Information retrieval in practice*, volume 520. Addison-Wesley Reading.

[Dantzig, 1957] Dantzig, G. B. (1957). Discrete-variable extremum problems. *Operations research*, 5(2):266–288.

[Davey and Priestley, 2002] Davey, B. A. and Priestley, H. A. (2002). *Introduction to lattices and order*. Cambridge university press.

[de Loof, 2010] de Loof, K. (2010). *Efficient computation of rank probabilities in posets*. PhD thesis, Universiteit Gent.

[Deb and Jain, 2013] Deb, K. and Jain, H. (2013). An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part i: solving problems with box constraints. *IEEE transactions on evolutionary computation*, 18(4):577–601.

[Deb et al., 2002] Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002). A fast and elitist multi-objective genetic algorithm: Nsga-ii. *IEEE transactions on evolutionary computation*, 6(2):182–197.

[Durante et al., 2013] Durante, F., Fernandez-Sanchez, J., and Sempi, C. (2013). A topological proof of sklar's theorem. *Applied Mathematics Letters*, 26(9):945–948.

[E.Knuth, 1997] E.Knuth, D. (1997). *The Art of Computer Programming*, volume 1. Addison-Wesley, Massachusetts, 3 edition.

[Halmos, 2013] Halmos, P. R. (2013). *Measure theory*, volume 18. Springer.

[Huber, 2006] Huber, M. (2006). Fast perfect sampling from linear extensions. *Discrete Mathematics*, 306(4):420–428.

[Ishibuchi et al., 2016] Ishibuchi, H., Imada, R., Setoguchi, Y., and Nojima, Y. (2016). Performance comparison of nsga-ii and nsga-iii on various many-objective test problems. In *2016 IEEE Congress on Evolutionary Computation (CEC)*, pages 3045–3052. IEEE.

[Jones, 1979] Jones, D. S. (1979). *Elementary information theory*. Clarendon Press.

[Kendall, 1955] Kendall, M. G. (1955). *Rank correlation methods*. Griffin, 2 edition.

[Kumara et al., 2020] Kumara, A., Wub, G., Alic, M., Luob, Q., Mallipeddid R., Suganthane, P. N., and Swagatam Das, S. (2020). Guidelines for real-world multi-objective constrained optimisation competition. https://raw.githubusercontent.com/P-N-Suganthan/2021-RW-MOP/main/Revised_Guideline_RWCMOP.pdf.

[Li et al., 2015] Li, B., Li, J., Tang, K., and Yao, X. (2015). Many-objective evolutionary algorithms: A survey. *ACM Computing Surveys (CSUR)*, 48(1):1–35.

[Liefooghe et al., 2009] Liefooghe, A., Jourdan, L., and Talbi, E.-G. (2009). A unified model for evolutionary multi-objective optimization and its implementation in a general purpose software framework. In *2009 IEEE Symposium on Computational Intelligence in Multi-Criteria Decision-Making (MCDM)*, pages 88–95. IEEE.

[Martello, 1990] Martello, S. (1990). Knapsack problems: algorithms and computer implementations. *Wiley-Interscience series in discrete mathematics and optimization*.

[Mitchell, 1998] Mitchell, M. (1998). *An introduction to genetic algorithms*. MIT press.

[Noman et al., 2011] Noman, N., Bollegala, D., and Iba, H. (2011). An adaptive differential evolution algorithm. In *2011 IEEE Congress of Evolutionary Computation (CEC)*, pages 2229–2236. IEEE.

[Palakonda et al., 2018] Palakonda, V., Ghorbanpour, S., and Mallipeddi, R. (2018). Pareto dominance-based moea with multiple ranking methods for many-objective optimization. In *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 958–964. IEEE.

[Resnick and Varian, 1997] Resnick, P. and Varian, H. R. (1997). Recommender systems. *Communications of the ACM*, 40(3):56–59.

[Reyes-Fernández-de Bulnes et al., 2019] Reyes-Fernández-de Bulnes, D., Bolufé-Röhler, A., Tamayo-Vera, D., et al. (2019). Multi-objective optimization approach based on minimum population search algorithm. *GECONTEC: Revista Internacional de Gestión del Conocimiento y la Tecnología*, 7(2):1–19.

[Shang et al., 2020] Shang, K., Ishibuchi, H., He, L., and Pang, L. M. (2020). A survey on the hypervolume indicator in evolutionary multiobjective optimization. *IEEE Transactions on Evolutionary Computation*, 25(1):1–20.

[Sklar, 1959] Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, 8:229–231.

[Zitzler and Thiele, 1999] Zitzler, E. and Thiele, L. (1999). Multiobjective evolutionary algorithms: a comparative case study and the strength pareto approach. *IEEE transactions on Evolutionary Computation*, 3(4):257–271.

Supplementary Materials

A   FURTHER EXAMPLES FOR SIMPLE RELATIONS AND MEASURES

In the first two supplementary examples, we consider a situation typical in economics or multi-objective optimization. Later, we show how the proposed concepts apply to arbitrary transitive relations.

A.1   Continuous Measures

Let us again consider $R_{\llcorner}$ as defined in Section 2.3. The relation $R_{\llcorner}$ models preference for small values of x_1 and x_2. However, instead of assuming X to be a finite subset of $\mathbb{R}^2$, we now study the unit square $[0,1] \times [0,1]$ with three continuous measures: the Lebesgue area measure $dx_1 dx_2$, as well as $2x_2 dx_1 dx_2$ and $4x_1 x_2 dx_1 dx_2$. In each case, the total measure of the unit square equals to one. The Lebesgue area measure represents elements with two uniformly distributed characteristics x_1 and x_2; $2x_2 dx_1 dx_2$ represents rarefaction of items having small values of x_2, whereas $4x_1 x_2 dx_1 dx_2$ represents rarefaction of items having small values of both x_1 and x_2.

For each of the three above cases we show in Fig. 8a the set of at least k-Pareto optimal elements of measure 0.1, which corresponds to selecting the 10 best percent. All three sets demonstrate the qualitative behaviours expected from sets delimited by indifference curves when the corresponding rarefaction occurs. Indeed, the curve corresponding to the uniform distribution and the Lebesgue area measure $dx_1 dx_2$ is symmetric. Also, in this case, $\mathrm{po}(x_1, x_1) = x_1 x_2$, and the sets of at least k-Pareto optimal elements are the sets situated below arcs of hyperbola defined by the equation $x_1 x_2 = k$, see Section 4.2 for more details. Applying rarefaction with respect to x_2 prioritises smaller values of this characteristic. This is represented by shifting upwards the right part of the hyperbola arc, see the curve for $2x_2 dx_1 dx_2$. Indeed, in this case, the small values of x_2 are observed less often. This results in selecting additional elements with large values of x_1 but relatively small values of x_2 to compensate for this rarefaction. Finally, rarefaction with respect to both x_1 and x_2 results in the fact that the small values of both characteristics are observed less often. Thus, elements with larger values of x_1 and x_2 should be selected to generate a selection of the required measure. It results in the shift of the hyperbola upwards following the direction of the main diagonal, see the curve for $4x_1 x_2 dx_1 dx_2$.
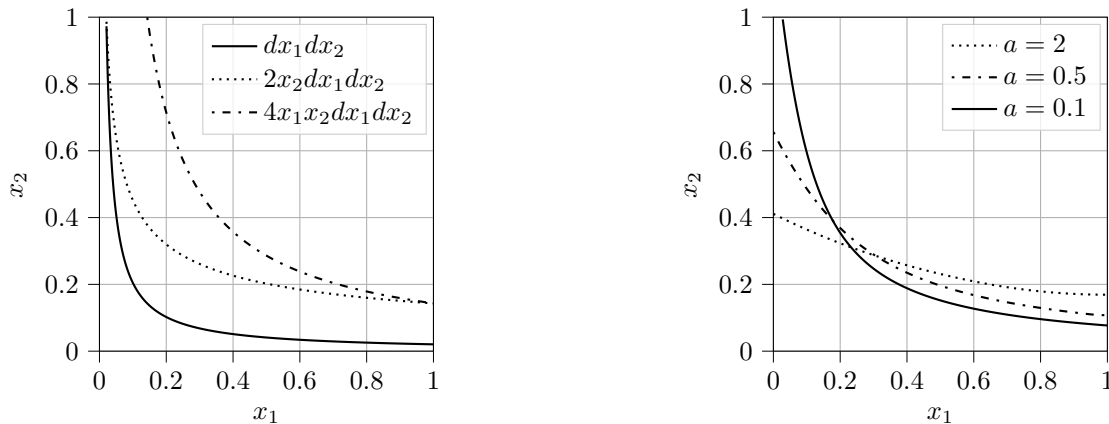
(a) At least k-Pareto optimal elements of measure 0.1 for the relation $R_{\llcorner}$ and three measures: the Lebesgue area measure $dx_1 dx_2$, rarefaction of x_2 defined by $2x_2 dx_1 dx_2$, and rarefaction of x_1 and x_2 defined by $4x_1 x_2 dx_1 dx_2$.

(b) At least k-Pareto optimal elements of measure 0.25 for the cone-based relations R_a and the Lebesgue area measure. Larger values of a represent higher maximum accepted trade-offs.

Figure 8: Examples at least k-Pareto optimal elements with two continuously distributed characteristics x_1, x_2.

A.2 Cone-based Relations

Let us again consider the unit square, the Lebesgue measure, and a positive constant a. However, this time the preference relation is defined as follows: yR_ax iff $y_2 \leq x_2$ and $x_2 - y_2 \geq a(y_1 - x_1)$. The above relation R_a is an example of a cone-based relation illustrated in Fig. 9. This relation has the intuitive meaning of giving up (x_1, x_2) for getting (y_1, y_2) if the improvement (diminution) in the second characteristic is at least a times the trade-off (increase) in the first characteristic. In this case, selections are sets delimited by descending curves $x_2 = f(x_1)$ such that $-\frac{1}{a} \leq \frac{df}{dx_1} \leq 0$.
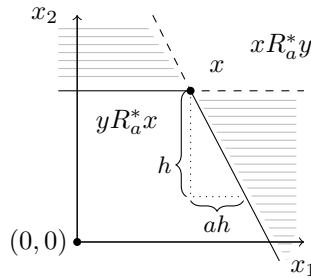

Figure 9: An illustration of a cone-based relation R_a.

Let us now consider the sets of at least k-Pareto optimal elements of measure 0.25 for the three values of a: $a = \frac{1}{10}$, $a = \frac{1}{2}$, and $a = 2$, see Fig. 8b. Larger values of a represent higher maximum accepted trade-offs. This is represented by the gradual degeneration of the hyperbola into a straight horizontal line when a increases. As shown in the figure, the three sets demonstrate plausible behavior. In the situation discussed in Appendix A.1, the relation $R_{\llcorner}$ corresponds to the extreme case of the relation R_a with $a = 0$.

A.3 Transitive Relations

In general, it is possible to show that if R^* is transitive, then for any k, the set T_k is a selection. In particular, if R is a partial order relation, μ is strictly positive, and X is countable, we obtain a linear extension of R when sorting X by increasing values of po and sorting ties in any order. Selections are represented by downsets. The latter are obtained when topologically sorting X and taking the first n elements, for any n. If μ is the counting measure, then po(x) is simply the number of elements that can be reached by following downwards the edges of the corresponding Hasse diagram. An example of such a relation represented by its Hasse diagram is depicted in Fig. 10.
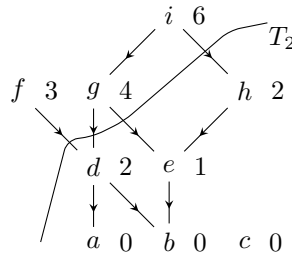

Figure 10: An illustration of simple partial order relation. The values of po are shown by numbers, and the set of at least 2-Pareto optimal elements T_2 is delimited by a curve.

B FURTHER THEORETICAL EXPLORATION

B.1 Efficient Computation of Choice

In general, computation of choice may be simplified by performing the change of variable $y = \mathrm{po}(x)$ in Eq. (1).

Proposition 1. *For any selection of at least k-Pareto optimal elements T_k such that $\mu(T_k) < +\infty$*

$$\mathrm{cho}(T_k) = \mu(T_k)^2 - 2\int_{[0,k]} x \, d(\mathrm{po} *\mu)(x), \tag{8}$$

where $\mathrm{po} *\mu$ is the image measure defined by $(\mathrm{po} *\mu)([a,b]) = \mu(\mathrm{po}^{-1}([a,b]))$.

B.2 Independence on Marginal Distribution

We study a probability space (Ω, Σ, P). We consider two random variables X_1 and X_2, as well as the partial order relation $R_{\Omega_\llcorner}$ defined on Ω as follows:

$$\omega_x R_{\Omega_\llcorner} \omega_y \text{ iff } X_1(\omega_x) \leq X_1(\omega_y) \text{ and } X_2(\omega_x) \leq X_2(\omega_y).$$

Let us consider X to be continuous. Sklar's theorem [Sklar, 1959, Durante et al., 2013] states that the cumulative distribution function $F(x_1, x_2)$ can be represented as $C(F_1(x_1), F_2(x_2))$ for a copula C. Marginas of X_1 and X_2 are fully described by the marginal cumulative probability distributions F_1 and F_2, whereas the copula describes the dependence structure between X_1 and X_2. The copula can be considered as a joint cumulative distribution function having two uniform marginal distributions on $[0,1]$. Below we show that the introduced concepts do not depend on the marginal distribution of X_1 and X_2.

Proposition 2. *For a continuous random vector (X_1, X_2) and the relation $R_{\Omega_\llcorner}$, k-Pareto optimality, choice, diversity and selections offering maximum choice only depend on the copula C of X_1, X_2.*

Proof. Let us consider the mapping

$$
\begin{array}{rccc}
G & : & \Omega & \to & [0,1]^2, \\
& & \omega_x & \mapsto & (F_1(x_1), F_2(x_2)).
\end{array}
$$

We consider $R_{\Omega_\llcorner}$ and P defined on Ω. At the same time, on $[0,1]^2$ we consider $R_\llcorner$ defined by $(x_1, x_2) R_\llcorner (y_1, y_2)$ iff $x_1 \leq x_2$ and $y_1 \leq y_2$, as well as the image measure $G *P$ defined on $[0,1]^2$ by $(G *P)(A) = P(G^{-1})(A)$. The map G preserves probabilities in the sense that for any measurable A in Ω we have $(G *P)(G(A)) = P(A)$. Moreover, G preserves the relations in the sense that $x R_{\Omega_\llcorner} y$ iff $G(x) R_\llcorner G(y)$. Selections are preserved in the sense that if S is a selection for $R_{\Omega_\llcorner}$, then $G(S)$ is a selection for $R_\llcorner$. Ignoring negligible subsets, this mapping between selections is one-to-one. Therefore, G also preserves selections, k-Pareto optimality, choice, and diversity.

The proposition finally results from the fact that $G *P$ only depends on the copula. This is a consequence of the fact that for any $(a_1, a_2) \in [0,1]^2$, we have $(G *P)([0,a_1] \times [0,a_2]) = C(a_1, a_2)$. This equality is a result of the following statements: 1) continuity which guarantees that a_1 and a_2 can be written as $F_1(x_1)$ and $F_2(x_2)$ for some appropriate x_1 and x_2; 2) the definition of image measure; 3) the fact that $G = (F_1, F_2) \circ (X_1, X_2)$; and 4) the equality $F(x_1, x_2) = C(F_1(x_1), F_2(x_2))$. □

Proposition 3. *If X_1 and X_2 are two continuous independent random variables, then for the relation $R_{\Omega_\llcorner}$*

$$P(T_k) = k - k\ln(k), \qquad \mathrm{cho}(T_k) = (k - k\ln(k))^2 - k^2\left(\frac{1}{2} - \ln k\right).$$

Proof. Let us consider the map G. The image measure $G *P$ induced by G on $[0,1]^2$ is the Lesbegue area measure. Independently of P, we have $G(T_k) = \{(x_1, x_2) \in [0,1]^2 | x_1 x_2 \leq k\}$. Integration for Eq. (8) yields

$$P(\mathrm{po}^{-1}(]-\infty, x])) = P(T_x) = \int_{(x_1, x_2) \in [0,1]^2} (G(T_x)) = x - x\ln(x).$$

Therefore, po $*P = -\ln(x)dx$ and Eq. (8) results in

$$\text{cho}(T_k) = \text{cho}(G(T_k)) = (k - k\ln(k))^2 - 2\int_0^k x(-ln(x))dx = (k - k\ln(k))^2 - k^2\left(1/2 - \ln k\right).$$

□

B.3 Sorting by k-Pareto Optimality versus Sorting by Pareto Fronts

To further illustrate the difference between sorting by Pareto fronts and k-Pareto optimality, we demonstrate the results of sorting the elements of a 2-dimensional grid for the relation $R_{\llcorner}$ and counting measure in Fig. 11. The numbers on the plots represent the equivalence class (front) to which a point was assigned by the relative sorting procedure. As we can see, sorting by Pareto fronts splits the points by straight lines, see Fig. 11a. At the same time, sorting by po results in splitting by hyperbola-like curves, see Fig. 11b. We can also notice, that extreme solutions[16] are valued more when sorting by po. Indeed, most of the non-extreme solutions are pushed to further equivalence classes, as compared to sorting by Pareto fronts. This characteristic of po-based sorting is also clearly visible in Fig. 12. Here we present selections of $\mu = 0.2$ for the same relation and the set X composed of a large number of points placed on a regular grid within the shaded area.

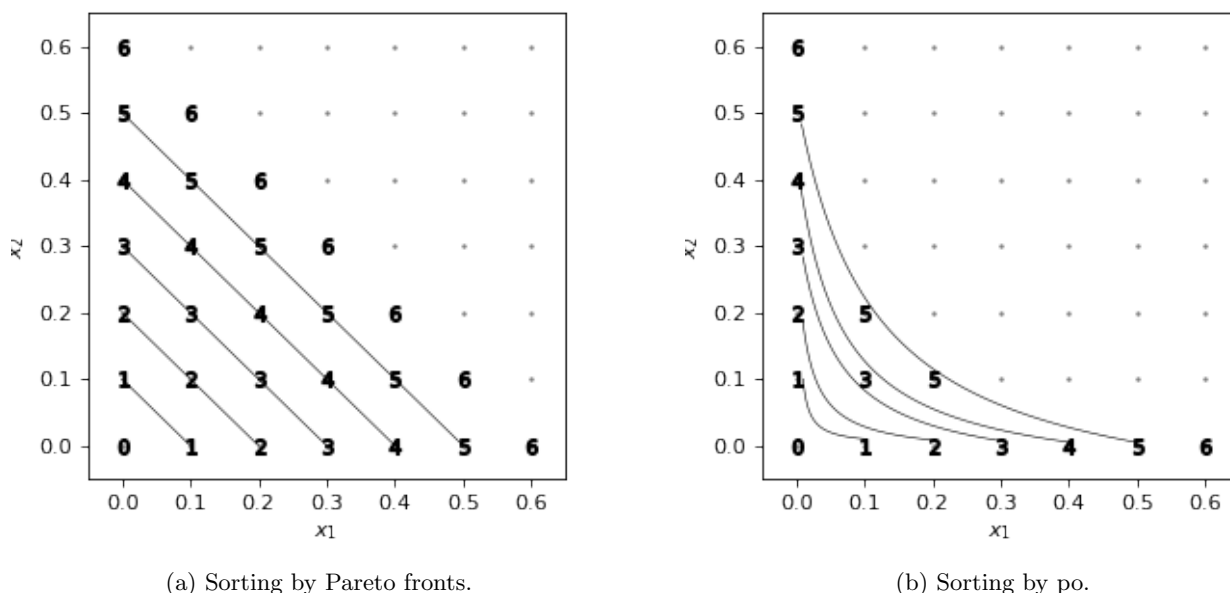(a) Sorting by Pareto fronts.

(b) Sorting by po.

Figure 11: Sorting points of a grid. The equivalence classes (fronts) are represented by numbers.

However, sorting by Pareto front does not always result in selections delimited by straight lines. Analysing the results of sorting for uniformly distributed points, we observe that both sorting methods result in hyperbola-like selections, see Fig. 13. This means that sorting by Pareto fronts is more sensitive to the topological structure of the analyzed space, while sorting by po preserves its characteristics.

We will now consider one more example. For any positive integer n and any α_n in $\{0,1\}$ one may consider the following deformation of a regular grid of n^2 points placed on the unit square:

$$A_n = \{(\frac{i}{n}, \frac{j}{n}) + \frac{\alpha_n}{2n^2}(j, i), \quad i \text{ and } j \text{ integers and } 0 \leq i, j < n\}.$$

On each of the of n^2 points one may place a weight of $1/n^2$. As n increases, for any α_n, the above distributions converge to the uniform distribution on the unit square. The selections obtained via sorting by k-Pareto

---

[16]*Extreme solution* here means that a solution is very good according to one criteria and is bad according to another.
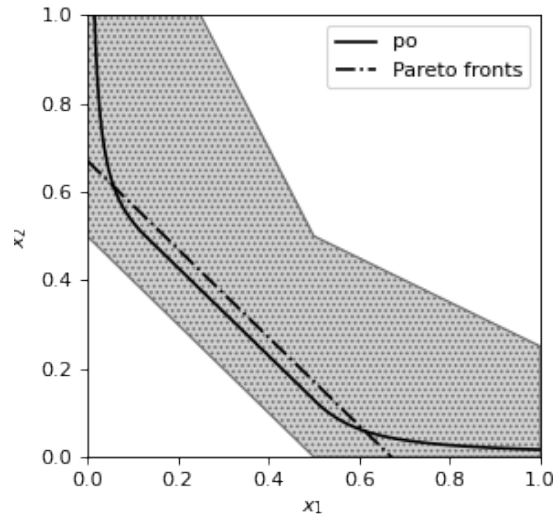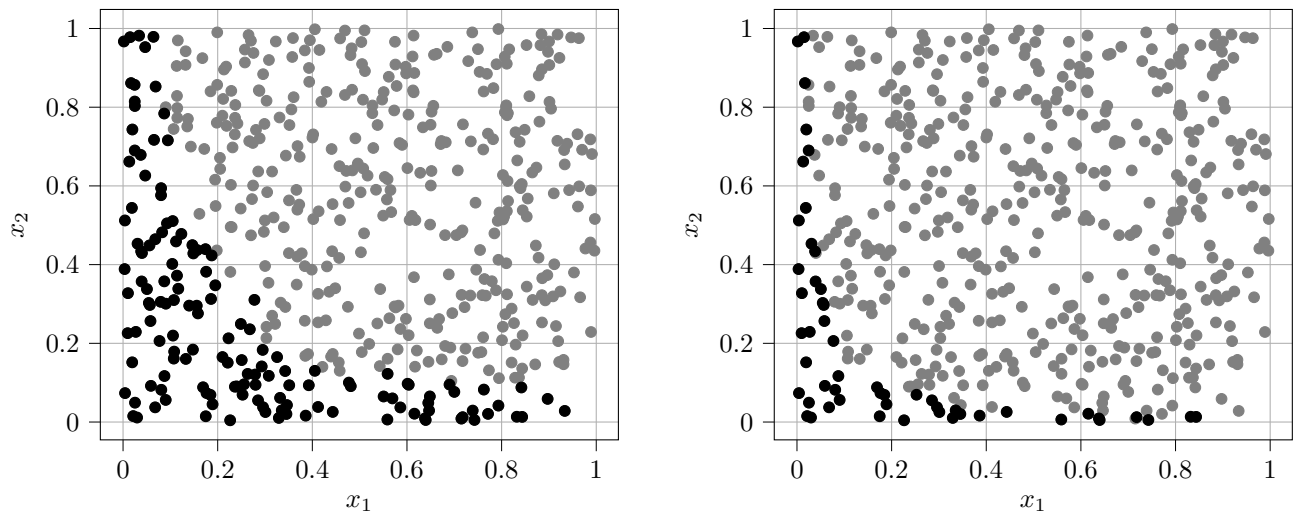
Figure 12: Selections of $\mu = 0.2$ for the set X composed of points in the shaded area.


(a) Sorting by Pareto fronts. The total number of equivalence classes (fronts) is 39.

(b) Sorting by po. The total number of equivalence classes (fronts) is 259.

Figure 13: Sorting 500 uniformly distributed points. Points in black belong to the first 10 equivalence classes (fronts). Note, that the total number of equivalence classes is larger for po-based sorting. The latter approach results in fewer ties.

optimality according to $R_{\llcorner}$ converge to the arcs of hyperbola obtained in Section 4.2. Choice converges as well. Formally, the distribution representing the points of A_n as well as the weights is $\sum_{x \in A_n} \delta_x / n$, where δ_x is the Dirac measure for the point x[17] and convergence means weak convergence [Billingsley, 1999, p. 7]. However, if $\alpha = -1$, then Pareto front sorting is the same as sorting according to the min function, whereas for $\alpha = 1$ it is the same as sorting according to the max function, see Fig. 14. Therefore, taking $\alpha_n = -1$ if n is even, and $\alpha_n = 1$ if n is odd makes the selections obtained via Pareto front sorting diverge when n tends to infinity, $n \to \infty$.

---

[17]δ_x is defined for a given $x \in X$ and any (measurable) set $A \subseteq X$ by $\delta_x(A) = 1$ if $x \in A$ and $\delta_x(A) = 0$ if $x \notin A$.
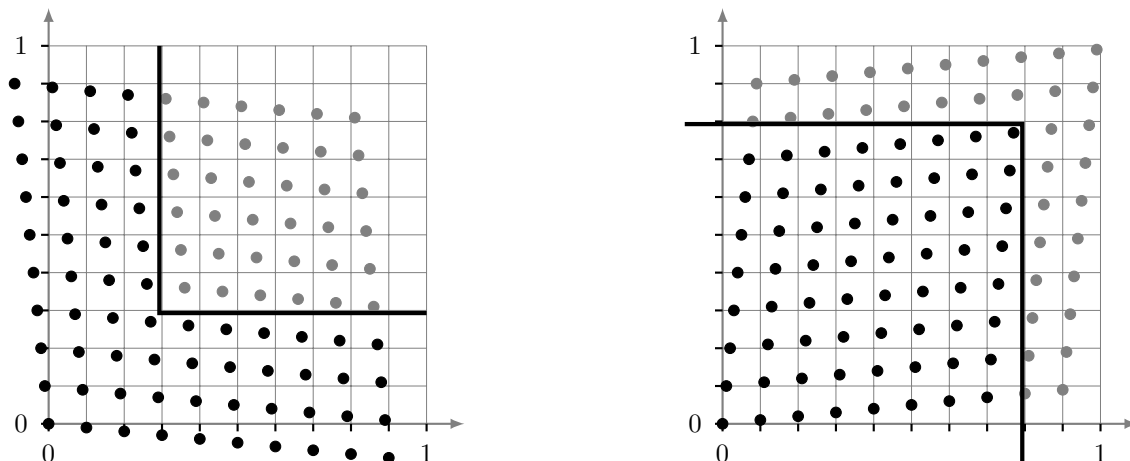
(a) If $\alpha_n = -1$, the selections are delimited by curves of equation $\min(x_1, x_2) = c$

(b) If $\alpha_n = 1$, the selections are delimited by curves of equation $\max(x_1, x_2) = c$

Figure 14: Selections of measure 0.64 (containing 64 points) obtained via sorting by Pareto fronts for deformations of a regular grid presented with the sets A_n, $n = 10$.

C   FURTHER PRACTICAL EXPLORATION

C.1   Additional Results for Genetic Optimization

In Section 5.2, we evaluate the performance of the genetic algorithms using the **hypervolume** indicator. In this section, we further analyze the behavior of both the state-of-the-art and the proposed algorithms with respect to other metrics. In particular, we study the fraction of solutions dominated by the solution of alternative algorithms and analyze the time complexity of the sorting procedure.

We calculate the percentage of **dominated solutions** as follows. For a given pair of algorithms *algorithm1* and *algorithm2*, we calculate how many solutions of *algorithm2* (*dominated algorithm*) are dominated by solutions of *algorithm1* (*dominating algorithm*). After that, we average the obtained results among all *dominating algorithms* to get an average fraction of dominated solutions, denoted by θ. Naturally, lower values of θ indicate better performance. We present the corresponding results in Fig. 15. We notice the following tendencies. *NSGA-II* and *PO-count* behave very similarly. For $n_k = 2$, the value of θ for these algorithms is around 20%. After that, it starts increasing and reaches its peak of approximately 45% for $n_k = 7$. Finally, it gradually decreases to 24% for $n_k = 25$. *NSGA-III* starts at a similar level and reaches its peak of approximately 30% for $n_k = 5$. After that, it decreases below 10% for $n_k = 7$ and stays relatively close to 0 for the larger numbers of knapsacks. These results demonstrate the superiority of *NSGA-III* over *NSGA-II* in the case of many-objective optimization. *PO-prob* starts at around 16%. However, for $n_k = 4$ the value of θ it already almost 0 and does not go up for larger numbers of knapsacks. This shows that the solutions produced by this algorithm are rarely dominated. Thereby, *PO-prob* is an effective approach for many-objective optimization problems.
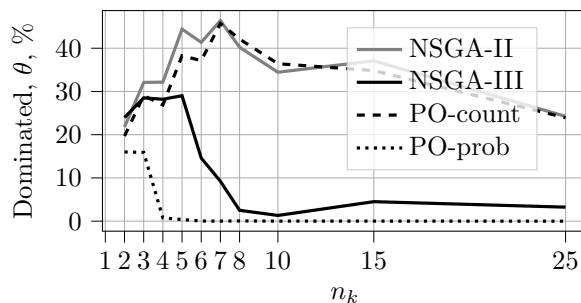

Figure 15: Average percentage of solutions dominated by other algorithms, θ.
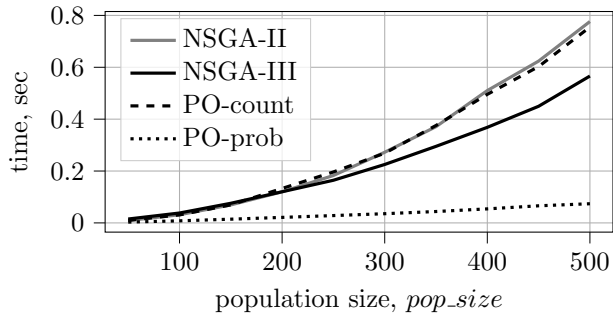
Figure 16: Sorting duration as a function of population size for 10 knapsacks, $n_k = 10$.

In Fig. 16, we demonstrate the dependence of sorting time on the population size for values of *pop_size* ranging from 50 to 500. The reported values are the averages over 100 independent executions of one iteration of the corresponding genetic algorithm. From the figure, we can see that *PO-prob* requires much less time than all other algorithm. The results for *NSGA-II* and *PO-count* tend to be very close, as in other experiments. This observation also has a theoretical explanation. Indeed, choosing the next generation for *NSGA-II* and *NSGA-III* has time complexity of $O(N^2M)$ and $max\{O(N^2M), O(N^2log_{M-2}N)\}$ respectively where M stands for number of objectives and N is the population size, see [Deb et al., 2002, Deb and Jain, 2013]. At the same time, sorting in *PO-prob* comes down to independent sorting procedures with respect to every objective. The time complexity of this procedure is $O(NMlog(N))$. These results are in line with the theoretical analysis presented in Section 5.1 and prove the computational efficiency of the approximate ranking calculation procedure used in *PO-prob*.

The maximum choice theorem (Theorem 1) has an intuitive interpretation in the context of genetic algorithms. Assume that the selection step is required to pick a selection of a given maximum size for breeding offspring, and both parents are chosen independently and at random form this selection. Then selections obtained via *k*-Pareto optimality-based sorting yield most offspring with parents offering choice. *Choice* here means that every parent is strictly superior to the other with respect to at least one objective, or both have the same values for all objectives.

C.2 Kendall's τ Rank Correlation Coefficient and Statistical Tests

Let us again consider the case of 2 continuous random variables introduced in Section 4.3. Let us assume that $X_2 = f(X_1)$ for some increasing function f. For almost all (x, y), either xRy or yRx holds. Thus $\mathrm{div}(\Omega) = 0$. Moreover, diversity only depends on the copula which encodes the dependency structure between X_1 and X_2, see Appendix B.2 and Section 4.1. Therefore, a value of $\mathrm{div}(\Omega)$ close to zero indicates X_1 and X_2 are strongly correlated via an increasing function. It leads to the idea that Kendall's τ rank correlation coefficient [Kendall, 1955] and diversity are strongly related concepts.

Let us consider a sample of n points $X = \{(x_{i1}, x_{i2})_{i \in \{1,2,\ldots,n\}}\}$. Duplicates almost never occur and the order in which points are drawn has no importance. Therefore, X should be treated like a set. We consider the counting measure $\#$ and the relation $R_{\llcorner}$ define on X. Diversity and choice of X are denoted by $\mathrm{div}_{\#}$ and $\mathrm{cho}_{\#}$.

Kendall's τ correlation coefficient is defined as follows

$$\tau = \frac{\#\,\mathrm{con} - \#\,\mathrm{dis}}{n_0},$$

where $\#\,\mathrm{con}$ is the number of concordant pairs (pairs that do not offer choice in our terminology), $\#\,\mathrm{dis}$ is the number of discordant pairs (pairs that do offer choice), and n_0 is the total number of pairs. As duplicates are discarded and the pairs are not ordered, $n_0 = n(n-1)/2$.

From the above remarks, we have $\mathrm{cho}_{\#} = 2\#\,\mathrm{dis} + n$ and $\#\,\mathrm{con} + \#\,\mathrm{dis} = n_0$. Combining these equalities, we obtain the following relation

$$\tau = \frac{n^2 + n - 2\,\mathrm{cho}_{\#}}{n^2 - n}.$$

Dividing the numerator and the denominator by n^2 and then neglecting $\frac{1}{n}$, we obtains that approximately

$$\tau \approx 1 - 2\operatorname{div}_\# .$$

This result means that the theory developed in this paper can be used for constructing non-parametric statistical tests generalizing Kendall's τ rank correlation coefficient and can be used for testing *partial correlation*. Below, we illustrate this property by building an indicator for distinguishing between wealthy and non-wealthy states.

A group of states might be considered *wealthy* if the following two conditions hold.

1. In the group there is no positive correlation between per capita income and the indicator representing education and health.

2. If a state belongs to a group of wealthy states, then all states having higher per capita income and better value of education and health indicator, must also belong to that group.

Now, we define the set of all wealthy states as the largest group of states that are wealthy. If Kendall's τ is used to compute correlation, and correlation is considered to be positive if $\operatorname{div}_\# < \frac{1}{2}$, then the elements of the above set can be easily identified.

Indeed, Corollary 1 says that the set of wealthy states must be a set of at-least k-Pareto optimal states for the relation *higher income* and *better education and health indicator*. For the year 2015[18], we took Gross National Income (GNI) per capita at purchasing power parity (PPP) as the *income indicator*, and the square root of the education and life expectancy as the *education and health indicator*[19]. The scatter plot in Fig. 17 shows the resulting division of states into wealthy and non-wealthy. We can observe that the wealthy states are defined as the states with GNI $\geq 20\,000\$$. This seems perfectly plausible.
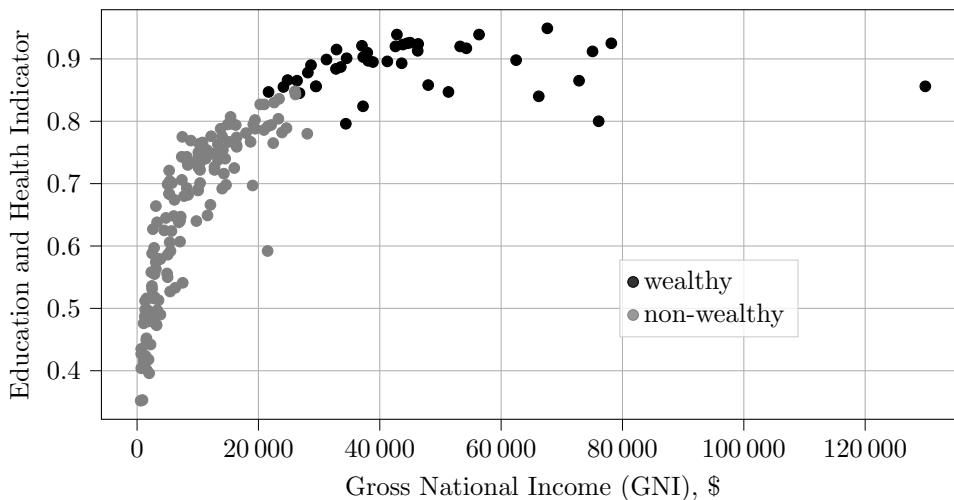

Figure 17: Separation between wealthy and non-wealthy states based on $\operatorname{div}_\#$.

C.3 Application to Recommender Systems

Let us again consider the housing example introduced in Section 2.1. If we aim to provide to the user a full set of possible alternative houses that might fit his preferences, then, according to Theorem 1, sorting available habitations by the increasing value of po is the best strategy. As it was discussed in Section 5.1, in the case of independent components of the underlying composite relations, the computation of po can be simplified by using tools from probability theory. Apart from computational efficiency, estimating po in this way has several additional advantages.

---

[18]Data source: United Nations Development Programme - Human Development Reports http://hdr.undp.org/en/data.

[19]As suggested by the human development index, see http://hdr.undp.org/sites/default/files/hdr2020_technical_notes.pdf.

- *Such sorting results in fewer ties and a meaningful score.* Indeed, sorting items x by increasing values of $\mathrm{po}(x)$, is the same as sorting by decreasing values of $-\log(\mathrm{po}(x))$. The self-information $-\log(F_i(x))$ [Jones, 1979], which is additive, indicates how much a characteristic i is valued. In this case, there is no need to introduce any arbitrary coefficients as it is done when sorting by a weighted mean of the characteristics x_i.

- If the condition of independence holds, then *rarer characteristics get valued more.* This makes sense from the economic point of view and is intuitively necessary for maximizing choice.

- If beyond the relation R, there is complete uncertainty about the user's complex needs, tastes and desires, then offering him a selection of maximum choice *maximizes the likelihood he finds an appropriate item.*

C.4 Constrained Multi-Objective Genetic Algorithms

C.4.1 Problem Definition

A Multiobjective Constrained Optimization Problem (CMOP) is a mathematical problem that is defined as follows [Kumara et al., 2020]:

Minimize
$$f_1(x), f_2(x), \ldots, f_M(x)$$

subject to

$$g_i(x) \leq 0, i \in \{1, 2, ..., ng\},$$
$$h_j = 0, j \in \{ng + 1, ng + 2, ..., ng + nh\},$$
$$L_k \leq x_k \leq U_k, h \in \{1, \ldots, D\},$$

where

- f_i represents the i-th objective function,

- M is the total number of conflicting objective functions,

- $x = (x_1, x_2, \ldots, x_D)$ is a solution vector of length D,

- L_k and U_k are the lower and upper bounds of the search space at the k-th dimension.

Numerically, we consider a constraint h_j to be verified iff $h_j \in [-\epsilon, \epsilon]$. A solution is feasible iff all $ng + nh$ constraints g_i and h_j are verified.

C.4.2 Problem Re-Definition with Preorder Relations

In a more general setting, we can represent a constraint $g_i \leq 0$ by the preorder relation R_{g_i} defined as follows:

$$xR_{g_i}y \text{ iff } \begin{cases} g_i(x) \leq 0 \text{ or} \\ g_i(x) \leq g_i(y). \end{cases} \tag{9}$$

And a constraint $h_j \in [a_j, b_j]$ can be represented by the preorder relation R_{h_j} defined as follows:

$$xR_{h_j}y \text{ iff } \begin{cases} h_j(x) \in [a_j, b_j] \text{ or} \\ h_j(y) \leq h_j(x) \leq a_j \text{ or} \\ b_j \leq h_j(x) \leq h_j(y). \end{cases} \tag{10}$$

Then, the combination of the constraints $g_i \leq 0, i \in \{1, 2, ..., ng\}$ and $h_j = 0, j \in \{ng + 1, ng + 2, ..., ng + nh\}$ can be represented by the preorder relation R_c defined as follows.

$$xR_cy \text{ iff } xR_{g_1}y \text{ and } \ldots \text{ and } xR_{g_{ng}}y \text{ and } xR_{h_{ng+1}}y \text{ and } \ldots \text{ and } xR_{h_{ng+nh}}y.$$

The objective consisting in minimizing f_i is represented by the preodrer relation R_{f_i}:

$$x R_{f_i} y \text{ iff } f_i(x) \leq f_i(y).$$

Minimization of all M objectives $f_1, \ldots, f_M$ is represented by the preorder relation R_f defined as follows

$$x R_f y \text{ iff } x R_{f_1} y \text{ and } \ldots \text{ and } x R_{f_M} y.$$

Thus, the above CMOP can be represented by the lexicographic preorder relation

$$x R_{cf} y \quad \text{iff} \quad x R_c^* y \text{ or } (x R_c^= y \text{ and } x R_f y),$$

where $x R_c^* y$ means "$x R_c y$ and not $y R_c x$", and $x R_c^= y$ means "$x R_c y$ and $y R_c x$". For the given R_{cf}, constrained Pareto optimal solutions [Kumara et al., 2020] are solutions that are not Pareto dominated by any other solution.

C.4.3 Solution

To solve the problem defined above, we can use the standard Adaptive Differential Evolution Algorithm jDE [Noman et al., 2011] with k-Pareto optimality for R_{cf} as a fitness function. For any point x, k-Pareto optimality of x is the likelihood a point drawn at random from the population strictly Pareto dominates x for R_{cf}. Smaller values of po mean better fitness. Under the independence assumption of objectives and constraints, we can easily compute k-Pareto optimality po(x). When saying $P(\{y|y R_{cf}^= x\}) = 0$, we assume the considered objectives and constraints are not constant on too large sets. Without this simplification, the computation becomes longer, see the derivation below.

$$
\begin{aligned}
\text{po}(x) &= P(\{y|y R_{cf}^* x\}), \\
&= P(\{y|y R_{cf} x\} - P(\{y|y R_{cf}^= x\}), \\
&= P(\{y|y R_{cf} x\} - 0, \\
&= P(\{y|y R_c^* x\} \cup \{y|y R_c^= x \text{ and } y R_f x\}), \\
&= P(\{y|y R_c^* x\}) + P(\{y|y R_c^= x \text{ and } y R_f x\}).
\end{aligned}
$$

Thus, if x satisfies all constraints, which means $x R_c^= (0, \ldots, 0, a_{ng+1}, \ldots, a_{ng+nh})$, then

$$
\begin{aligned}
\text{po}(x) &= P(\{y|y R_c^= x \text{ and } y R_f x\}), \\
&= P(\{y|y R_c^= (0, \ldots, 0, a_{ng+1}, \ldots, a_{ng+nh})\}) P(\{y|y R_f x\}).
\end{aligned}
$$

Now, let $F_i(z) = P(\{y|f_i(y) \leq z\})$ be the cumulative probability distribution of f_i. Then,

$$
\begin{aligned}
P(\{y|y R_f x\}) &= P\left(\bigcap_{i \in \{1, \ldots, M\}} \{y|f_i(y) \leq f_i(x)\} \right), \\
&= \prod_{i \in \{1, \ldots, M\}} P(y|f_i(y) \leq f_i(x)). \\
&= \prod_{i \in \{1, \ldots, M\}} F_i(f_i(x)).
\end{aligned}
$$

Otherwise, if at least one constraint is not satisfied by x, then $P(\{y|y R_c^= x\}) = 0$ and

$$
\begin{aligned}
\text{po}(x) &= P(\{y|y R_c^* x\}) \\
&= P(\{y|y R_c x\}) \\
&= \prod_{i \in \{1, \ldots, ng\}} G_i(g(x)) \prod_{j \in \{ng+1, \ldots, ng+nh\}} P(y|y R_{h_j} x).
\end{aligned}
$$

Any cumulative probability distribution defined above, F_i for f_i and G_i for g_i, can be estimated via its empirical cumulative probability distribution. Note, for a population $\{x_1, \ldots, x_k, \ldots, x_{ps}\}$ of size ps, and for any real valued function f, the empirical cumulative probability distribution $\hat{F}$ of f is defined as follows:

$$\hat{F}(z) = \frac{\#(\{x_k | f(x_k) \leq z\})}{ps}.$$

In a similar way, $P(y | y R_{h_j} x)$ can be estimated by $\widehat{H_j}^*(h_j(x))$, where $\widehat{H_j}^*(z)$ is defined as

$$\widehat{H_j}^*(z) = \begin{cases} \frac{\#(\{x_k | z \leq h_j(x_k) \leq b_j\}}{ps} & \text{if } z < a_j, \\ \frac{\#(\{x_k | a_j \leq h_j(x_k) \leq b_j]\}}{ps} & \text{if } a_j \leq z \leq b_j, \\ \frac{\#(\{x_k | a_j \leq h_j(x_k) \leq z\}}{ps} & \text{if } z > b_j. \end{cases} \tag{11}$$

The computation of every $\hat{F}_i$ can be performed as follows:

- sort the values $f(x_k)$ in increasing order, and store them in an array;

- create two new arrays;

- loop over the sorted values $f(x_k)$; each time a new distinct value $f(x_k)$ is encountered:
  - append the previously encountered $f(x_k)$ to the first array,
  - append to the second array the loop counter, which is equal to the value of the empirical cumulative probability distribution $\hat{F}_i$ of the previously encountered value $f(x_k)$.

Thus, retrieving $\hat{F}_i(x)$ can be performed via a binary lookup with run time $O(\log ps)$. Computation of $\widehat{H_j}^*$ can be performed in the same way. In this case, all three cases of the definition in Eq. (11) are treated separately. Moreover, we have the estimation

$$P(\{y | y R_c^=(0, \ldots, 0, a_{ng+1}, \ldots, a_{ng+nh})\}) = \prod_{i \in \{1, \ldots, ng\}} \hat{G}_i(0) \prod_{j \in \{ng+1, \ldots, ng+nh\}} \widehat{H_j}^*(a_j).$$

Finally, it is possible to show that the total run time of the k-Pareto optimality based sorting is $O((ng + nh + M)ps \log ps)$.

C.5 Exploratory Database Queries

Simple database queries q, objectives, and constraints in optimization problems often consist in requiring a continuous attribute to be in a given interval, or a discrete attribute to be equal to a given value. Conceptually, those queries are boolean functions. Complex queries are often conjunctions of the form $r = q_1 \wedge q_2 \wedge \cdots \wedge q_n$.

In our formalism, these simple queries translate into simple pre-order relations of the form $x R y$. Requiring an element to be in an interval can be represented by $x R_q y$ iff x is in the desired interval, or x is not situated further from the interval than y[20]. Requiring an attribute to be equal to a given value translates into the relation $x R_{q'} y$ if for x the attribute takes the required value[21]. Complex queries then translate into the composite relations of the form $R_r = R_{q_1} \wedge R_{q_2} \wedge \cdots \wedge R_{q_n}$. The simple sub-relations R_{q_2} are pre-order relations, and, therefore, R_r is also a pre-order relation and is transitive. However, these relations are not partial order relations, as reflexivity does not necessarily hold. A "topological" sorting according to our partial order relation R_r can be viewed as a valid fuzzy relaxation of the strict functional query. There are many possible fuzzy relaxations and the problem is to find one that is suitable for a given application. The k-Pareto optimality is one of such fuzzy extensions of the query. It is 0 if all criteria are satisfied, and higher values of k-Pareto optimality indicate worse results. The maximum choice theorem applies here, and the user is offered the maximum choice. This is of particular

---

[20]Hence, the strict version of this relation is defined as follows: $x R_q^* y$ iff x is in the desired interval and y is not, or if x is situated closer to the desired interval.

[21]The strict version of this relation is defined as $x R_{q'}^* y$ if for x the attribute takes the required value but not for y.

interest for exploratory queries, such as job search, especially, if there are no items in the database that satisfy all the criteria. Direct brute-force search for selections offering maximum choice is unfeasible as there are too many selections to consider.

Moreover, in the above formalism, one can treat classical optimization objectives in the same way. Maximizing an attribute x can be represented by the relation xRy iff $x \geq y$, and the minimization can be represented by the relation $\leq$. In the above framework, negation can be represented via the relation R^{-1} defined by $xR^{-1}y$ iff yRx.

C.6   Scheduling Algorithms

In the case of scheduling algorithms, xRy can be given the meaning 'x depends on y'. Then, selections represent sets of tasks that remain to be processed. Having a large choice means *having much freedom to parallelize tasks* or *having flexibility in case the rescheduling is required.*