CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Scaling-invariant maximum margin preference learning
Authors	Montazery, Mojtaba;Wilson, Nic
Publication date	2020-10-19
Original Citation	Montazery, M. and Wilson, N. (2021) 'Scaling-invariant maximum margin preference learning', International Journal of Approximate Reasoning, 128, pp. 69-101. doi: 10.1016/j.ijar.2020.10.006
Type of publication	Article (peer-reviewed)
Link to publisher's version	https://www.sciencedirect.com/science/article/pii/S0888613X20302401 - 10.1016/j.ijar.2020.10.006
Rights	© 2020 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/) - http://creativecommons.org/licenses/by/4.0/
Download date	2026-11-08 10:45:37
Item downloaded from	https://hdl.handle.net/10468/10802

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

International Journal of Approximate Reasoning 128 (2021) 69–101

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Scaling-invariant maximum margin preference learning

Mojtaba Montazery, Nic Wilson *

Insight Centre for Data Analytics, School of Computer Science and IT, University College Cork, Ireland

ARTICLE INFO

Article history:
Received 29 September 2019
Received in revised form 24 September 2020
Accepted 7 October 2020
Available online 19 October 2020

Keywords:
Preference learning
Preference inference

ABSTRACT

One natural way to express preferences over items is to represent them in the form of pairwise comparisons, from which a model is learned in order to predict further preferences. In this setting, if an item a is preferred to the item b, then it is natural to consider that the preference still holds after multiplying both vectors by a positive scalar (e.g., $2a \succ 2b$). Such invariance to scaling is satisfied in maximum margin learning approaches for pairs of test vectors, but not for the preference input pairs, i.e., scaling the inputs in a different way could result in a different preference relation being learned. In addition to the scaling of preference inputs, maximum margin methods are also sensitive to the way used for normalizing (scaling) the features, which is an essential pre-processing phase for these methods. In this paper, we define and analyse more cautious preference relations that are invariant to the scaling of features, or preference inputs, or both simultaneously; this leads to computational methods for testing dominance with respect to the induced relations, and for generating optimal solutions (i.e., best items) among a set of alternatives. In our experiments, we compare the relations and their associated optimality sets based on their decisiveness, computation time and cardinality of the optimal set.
© 2020 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

There is a growing trend towards personalisation for services in many real-world application domains, such as e-commerce, marketing, and entertainment. This involves capturing user preferences over alternative choices, e.g., products, movies and hotels. One may view this as an enhanced variation of supervised learning, known as *preference learning*, where instead of tagging an instance with a single label, preference relations are expressed over instances [1,2]. One natural way to express preferences over items is to represent them in the form of pairwise comparisons, stating that one alternative a is preferred over another one b, where an alternative is associated with a feature vector, i.e., a vector of values for a number of features.

An established approach for modelling preferences makes use of the concept of a *utility function* that is learned from preference input pairs. Then, for a pair of test vectors (α, β), this function assigns an abstract degree of utility to each test vector, implying which test vector is preferred to which [3]. Support Vector Machine (SVM) approaches [4–6] have inspired the development of several methods for learning the utility function, such as OrderSVM [7], SVOR [8] and SVMRank [9].

In a method such as SVMRank, when the utility function has been learned, rescaling a pair of test vectors makes no difference to the result, i.e., α is preferred to β if and only if $r\alpha$ is preferred to $r\beta$ for any strictly positive scale factor r. The same does not hold for the input pairs: different ways of scaling preference input pairs may lead to a very different

* Corresponding author.
E-mail addresses: mojtaba.montazery@insight-centre.org (M. Montazery), nic.wilson@insight-centre.org (N. Wilson).

https://doi.org/10.1016/j.ijar.2020.10.006
0888-613X/© 2020 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

utility function being learned. However, it is arguable that in many contexts, a preference for a over b can be considered as conveying essentially equivalent information to a preference for ra over rb. For instance, knowing that the movie with the feature vector a is preferred to one with the feature vector b, we would often expect that $2a$ is preferred to $2b$. This suggests defining a more cautious preference relation by saying that a test vector α is *inputs-scaling-invariant preferred* to β if α is preferred to β for all choices of scalings of preference input pairs.

An analogous form of preference relation considers the scaling of features, where α is *features-scaling-invariant preferred* to β if α is preferred to β for all choices of scalings of features. Part of the motivation for this is that for any SVM-based method is necessary to scale (normalize) features beforehand. This is because these methods are not invariant to the scale of their input feature spaces; for example, a particular feature with very large values, compared with the other features, might effectively veto the effect of other features on the objective function. Therefore, these methods are clearly sensitive to the way features are scaled [10,11]. The common practice for scaling is based on the properties of input instances [12–14]; as an example of a scaling method, the value of a feature is subtracted by the minimum of all values of that feature in the dataset and divided by the difference between the maximum and minimum. So, the scaling, and therefore the resulting preference relation, can sometimes depend strongly on precisely which preference inputs are received. There can thus be subjective, and even rather arbitrary, choices in the scaling of the feature spaces; different ways lead to different preference relations.

Taking into account both forms of rescaling mentioned above, we can also consider a still more cautious relation in which α is *scaling-invariant* preferred to β if it is preferred for all choices of scalings of features and preference input pairs simultaneously.

Other preference reasoning techniques based on a family of utility functions include e.g., [15,16]. Another technique instead of learning a utility function, is learning a binary preference relation that compares pairs of alternatives and determines which one dominates which. Some early works of this form are considered in [17,18]. Other forms of preference inference that involves some predefined assumptions about the structure of the preference relation are based on more qualitative models [19], or lexicographic structure [20–23], or minimising regret [24–26], or additive independent family of methods [27–30], or GAI [31–33,25,26,34,35], or UTA [36–39], or Choquet integral [40–42]. Moreover, CP-nets [43] could be used to learn and represent more specific and customized assumptions; see e.g., [44].

We focus in this work on linear preference models. In particular, we can choose a vector w of weights, one weight for each feature, and consider the relation $\succeq_w$ given by: $\alpha \succeq_w \beta$ if and only if $w \cdot \alpha \geq w \cdot \beta$, where $w \cdot \alpha$ means $\sum_{i-1}^{n} w(i)\alpha_i$, and α and β are n-dimensional feature vectors representing alternatives. The vector w is chosen so that $\succeq_w$ is consistent with the preference inputs (which we assume are consistent themselves).

The consistency-based relation is given by: α is preferred to β if and only if it is preferred for all $\succeq_w$ with w consistent with the preference inputs. This can lead to a rather weak relation, leading to relatively uninformative guidance for a decision maker. In contrast the maximum margin relation generates a total pre-order that is consistent with the preference inputs Λ, by choosing a single vector w, which we call $\omega^\star_\Lambda$. This vector is chosen to maximise the degree of satisfaction with the preference inputs, corresponding to maximising the margin, similarly to SVM methods (for the case of consistent inputs).

Our novel preference relations are based on extending the maximum margin preference relation, by considering different forms of scaling. First we consider rescaling the preference inputs, generating a preference relation that is invariant to scaling up or down one or more of the preference inputs; for example, the preference relation does not change if we replace an input preference of $\alpha > \beta$ by $2\alpha > 2\beta$. This preference relation is constructed by considering all possible ways of rescaling the preference inputs Λ, which leads to a set $\mathrm{SI}(\Lambda)$ of vectors w, and the associated preference relation $\succcurlyeq^{\mathrm{I}}_\Lambda$ is given by: $\alpha \succcurlyeq^{\mathrm{I}}_\Lambda \beta$ if and only if $\alpha \succeq_w \beta$ for all $w \in \mathrm{SI}(\Lambda)$. Each element of $\mathrm{SI}(\Lambda)$ corresponds to the maximum margin solution w with respect some rescaling of the preference inputs. We derive a simple characterisation of this relation, which allows dominance to be efficiently checked.

We take a similar approach with rescaling the domains of the features. As mentioned above, a linear change of scale of any domain can significantly change the result for SVM methods and for the maximum margin preference relation. This is unfortunate because the choice of feature domains can be somewhat arbitrary. We modify the max-margin relation to form a relation $\succcurlyeq^{\mathrm{F}}_\Lambda$ that is invariant with respect to rescaling of the feature domains. This relation can be expressed in terms of a set $\mathrm{SF}(\Lambda)$, where $\alpha \succcurlyeq^{\mathrm{F}}_\Lambda \beta$ if and only if $\alpha \succeq_w \beta$ for all $w \in \mathrm{SF}(\Lambda)$; essentially, each w in $\mathrm{SF}(\Lambda)$ corresponds with a max-margin solution based on rescaled domains. $\mathrm{SF}(\Lambda)$ is a more complex set than $\mathrm{SI}(\Lambda)$, but we characterise it a number of ways, leading again to a computational technique for dominance with respect to the preference relation $\succcurlyeq^{\mathrm{F}}_\Lambda$.

Rescaling of preference inputs will almost always change the relation, i.e., we almost always have $\succcurlyeq^{\mathrm{I}}_\Lambda$ different from $\succeq_{\omega^\star_\Lambda}$. However, for certain inputs Λ it can happen that $\succcurlyeq^{\mathrm{F}}_\Lambda$ equals $\succeq_{\omega^\star_\Lambda}$, i.e., the rescaling of features makes no difference. We characterise these situations, and develop a computational approach for testing this, i.e., in which situations the max-margin preference relation is robust with respect to rescaling of feature domains.

The natural next step is to generate a more robust relation still, that is invariant to both these kinds of rescaling, of preference inputs and of feature domains. This is defined in a similar way to the two previous relations: $\alpha \succcurlyeq^{\mathrm{I,F}}_\Lambda \beta$ if and only if $\alpha \succeq_w \beta$ for all $w \in \mathrm{SIF}(\Lambda)$, where each element of $\mathrm{SIF}(\Lambda)$ is the max-margin w for some rescaled version of the problem. Again, we characterise this dominance relation to enable computation.

We examine the relationships between these relations, and analyse two natural ways of defining the best solutions among an input, with respect to each of these relations. As the relations are based on the assumption that the input pairs are consistent, we briefly discuss three possible approaches to generate a consistent preference input set.

Our experimental testing involves derivatives of two real databases; the experiments compare the different relations based on (a) the number of test pairs in which one dominates the other; (b) the number of optimal solutions found according to the defined optimality operators; and (c) the computation time.

Summary of contributions It is clearly important that the information in decision support systems is reliable and trustworthy. Although SVM-based (maximum margin) approaches to learning preferences are attractive and well-founded, they can also be very sensitive to the form of the preference inputs. For instance, the choice of feature domains can be somewhat arbitrary, but can significantly affect the result. The paper considers different forms of robustness for preference learning; in particular, we developed three novel robust preference learning techniques; an approach that is invariant to the rescaling of the features' domains; one that is invariant to rescaling of the input vectors; and a method that expresses both kinds of invariance. For each of these, we develop characterisations that lead to computational methods. We also have developed a computational approach for testing when maximum margin preference relation is invariant to feature domain rescaling. We analyse relationships between the different forms of preference relations, and, based on the computational characterisations, we implemented and tested them, comparing the relative numbers of optimal solutions, according to two natural kinds of optimality. We demonstrate that the methods are all different, and that they can be feasible computationally, and do not necessarily lead to large sets of optimal solutions.

The rest of the paper is organised as follows. The next section introduces the terminology being used throughout the paper and explains two preliminary preference relations, namely the consistency-based relation and the maximum margin relation. Section 3 considers the effect of rescaling of preference input pairs, and characterises a preference relation that is invariant to the scaling of preference input pairs. Similarly, the two other relations, where features are rescaled and where both features and preference inputs are rescaled simultaneously, are characterised in Sections 4 and 5 respectively. We discuss three possible approaches to deal with inconsistencies in Section 6. The characterisations of relations lead to the computational methods in Section 7 for testing dominance with respect to the induced relations. In Section 8, we consider two kinds of optimality operator to choose a subset of alternatives as optimal solutions with regard to each preference approach. We report the experimental results in Section 9; Section 10 concludes, with a discussion of potential extensions. The appendix includes all the proofs of the formal results that are not included in the main body of the paper.

This work includes and extends work in two conference papers [45,46].

2. Preliminaries

In this section, we describe some notation and two preference relations that provide a basis for the following sections. Since there are inevitably many symbols and results to keep track of, Table 1 includes a glossary of symbols.

We assume that some user has told us that she prefers feature vector $a_i \in \mathbb{R}^n$ over $b_i \in \mathbb{R}^n$, for each $i \in I = \{1, \ldots, m\}$. Each tuple a_i or b_i in $\mathbb{R}^n$ represents an alternative that is characterised by n features, with $a_i(k)$ being the score for alternative a_i regarding the kth feature..[1] By assuming a linear weighting model, each pair (a_i, b_i) expresses a linear restriction $a_i \cdot w > b_i \cdot w$ on an unknown weight vector $w \in \mathbb{R}^n$ (the dot product $a_i \cdot w$ is equal to $\sum_{j=1}^{n} a_i(j)w(j)$). This linear weighting assumption is less restrictive than it sounds; for instance, we could form additional features representing e.g., pairwise products of the basic features, enabling a richer representation of the utility function.

We define Λ, the *preference inputs*, to be $\{\lambda_i : i \in I\}$, where for each i, $\lambda_i = a_i - b_i$. Then, a *feasible* w satisfies $\lambda \cdot w > 0$ for all $\lambda \in \Lambda$ (because $a_i \cdot w > b_i \cdot w$). Let us denote the feasible set by $\Lambda^>$ (=$\{w \in \mathbb{R}^n : \forall \lambda \in \Lambda, w \cdot \lambda > 0\}$), and associate the hyperplane $H_w = \{x \in \mathbb{R}^n : x \cdot w = 0\}$ with a feasible $w \in \Lambda^>$. Clearly, any feasible hyperplane contains the origin, and all $\lambda \in \Lambda$ are in the associated positive open half-space of the hyperplane. We also will almost always be assuming that the preference inputs are consistent, so that $\Lambda^>$ is non-empty. Later, in Section 6, we discuss how to cope with inconsistency in preference inputs.

We sum up some of these key notions in the following definition.

Definition 1 *($\Lambda^>$ and consistent Λ).* A set of preference inputs Λ is a (finite) subset of $\mathbb{R}^n$. The feasible set $\Lambda^> = \{w \in \mathbb{R}^n : \forall \lambda \in \Lambda, w \cdot \lambda > 0\}$ consists of all elements of $\mathbb{R}^n$ that are feasible (with respect to Λ). Set of preference inputs Λ is said to be consistent if $\Lambda^> \neq \emptyset$.

Example 1. Suppose that $n = 2$ and let the preference inputs Λ be $\{(2,1), (1,2), (1,1)\}$ (see Fig. 1(a)). Then, a feasible $w \in \mathbb{R}^2$ satisfies these three conditions: (i) $2w(1) + w(2) > 0$, (ii) $w(1) + 2w(2) > 0$ and (iii) $w(1) + w(2) > 0$. The feasible

[1] For ordinal features (e.g., a feature with {*Cold*, *Mild*, *Hot*} variables) each value can be replaced by a number, maintaining the order of values. For categorical features, one might use the one-hot encoding (a.k.a. 1-of-k coding scheme) to convert a feature with k categories to k Boolean features. For example a feature, that represents the type of car with values {*Sedan*, *SUV*, *Hatchback*}, is converted to three binary features: *Is_Sedan*, *Is_SUV*, and *Is_Hatchback*. Clearly, among these three features exactly one of them is true and the two others are false.

Table 1
The glossary of symbols being used throughout the paper.

Symbol	Meaning
n	number of features.
m	number of preference input pairs.
(a_i, b_i)	a preference input pair when a_i has been preferred to b_i.
I	defined as $\{1, \ldots, m\}$; i.e., the index set for input pairs.
$\cdot$	the dot product, e.g., $(2,3)\cdot(3,1) = 9$.
Λ	is the (finite) set of preference inputs; i.e., $\{\lambda_i : \forall i \in I, \lambda_i = a_i - b_i\}$; e.g., the black points marked in Fig. 1(a).
$\Lambda^{>}$	defined as $\{w \in \mathbb{R}^n : \forall \lambda \in \Lambda, w \cdot \lambda > 0\}$; e.g., convex open space above and to the right of the dotted lines, in Fig. 1(b).
$\Lambda^{\geq}$	defined as $\{w \in \mathbb{R}^n : \forall \lambda \in \Lambda, w \cdot \lambda \geq 1\}$; e.g., the union of the two shaded regions in Fig. 1(b).
Λ^{*}	defined as $\{w \in \mathbb{R}^n : \forall \lambda \in \Lambda, w \cdot \lambda \geq 0\}$; e.g., the closed convex space surrounded by dotted lines in Fig. 1(b).
$co(\Lambda)$	the smallest convex cone containing Λ; e.g., the darkly shaded region in Fig. 1(a).
$\succcurlyeq_{\Lambda}^{C}$	the consistency-based relation; i.e., $\alpha \succcurlyeq_{\Lambda}^{C} \beta$ iff for all $w \in \Lambda^{>}$, $w \cdot (\alpha - \beta) \geq 0$.
$S - \{x\}$	from the set S, the element x is excluded.
$\Lambda^{*} + \{u\}$	for the vector $u \in \mathbb{R}^n$, defined as $\{w + u : w \in \Lambda^{*}\}$.
$\|w\|$	Euclidean norm of w.
$\omega_{\Lambda}^{\star}$	the element with minimal norm in $\Lambda^{\geq}$; e.g., $(0.5, 0.5)$ in Fig. 1(b).
$\succcurlyeq_{\Lambda}^{mm}$	the maximum margin preference relation; i.e., $\alpha \succcurlyeq_{\Lambda}^{mm} \beta$ iff $\alpha \cdot \omega_{\Lambda}^{\star} \geq \beta \cdot \omega_{\Lambda}^{\star}$.
$\mathbb{R}_{+}$	the set of strictly positive reals.
$\mathbb{R}_{+}^{m}$	the set of vectors $u \in \mathbb{R}^m$ with each component strictly positive.
$\mathbf{t}$	$\in \mathbb{R}_{+}^{m}$, the rescaling vector for preference inputs.
$\Lambda_{\mathbf{t}}$	defined as $\{\mathbf{t}(i)\lambda_i : \forall i \in I\}$, i.e., preference inputs being rescaled by $\mathbf{t}$.
$\succcurlyeq_{\Lambda}^{I}$	the relation that is invariant to the rescaling of inputs; i.e., $\alpha \succcurlyeq_{\Lambda}^{I} \beta$ iff for all $\mathbf{t} \in \mathbb{R}_{+}^{m}$, $\alpha \succcurlyeq_{\Lambda_{\mathbf{t}}}^{mm} \beta$.
$\mathrm{SI}(\Lambda)$	defined as $\{\omega_{\Lambda_{\mathbf{t}}}^{\star} : \mathbf{t} \in (0,1]^m\}$; e.g., the darkly shaded region in Fig. 1(b).
τ	$\in \mathbb{R}_{+}^{n}$, the rescaling vector for features.
τ^{-1}	$\in \mathbb{R}_{+}^{n}$, given by $\tau^{-1}(j) = 1/\tau(j)$ for all $j \in \{1, \ldots, n\}$.
$\odot$	pointwise multiplication, e.g., $(2,3)\odot(3,1) = (6,3)$.
$\Lambda \odot \tau$	defined as $\{\lambda \odot \tau : \forall \lambda \in \Lambda\}$, i.e., features being rescaled by τ.
$\succcurlyeq_{\Lambda}^{F}$	the relation that is invariant to the rescaling of features; i.e., $\alpha \succcurlyeq_{\Lambda}^{F} \beta$ iff for all $\tau \in \mathbb{R}_{+}^{n}$, $\alpha \odot \tau \succcurlyeq_{\Lambda\odot\tau}^{mm} \beta \odot \tau$.
$\mathrm{SF}(\Lambda)$	defined as $\{\omega_{\Lambda\odot\tau}^{\star} \odot \tau : \tau \in \mathbb{R}_{+}^{n}\}$; e.g., the part of the line segment $x + y = 1$ strictly within the first quadrant in Fig. 1(b).
$\succcurlyeq_{\Lambda}^{I,F}$	the relation that is invariant to the rescaling of features and inputs simultaneously.
$\mathrm{SIF}(\Lambda)$	defined as $\{\omega_{\Lambda_{\mathbf{t}}\odot\tau}^{\star} \odot \tau : \mathbf{t} \in (0,1]^m, \tau \in \mathbb{R}_{+}^{n}\}$; e.g., the part of the shaded regions that is strictly within the first quadrant (so not including the axes) in Fig. 1(b).

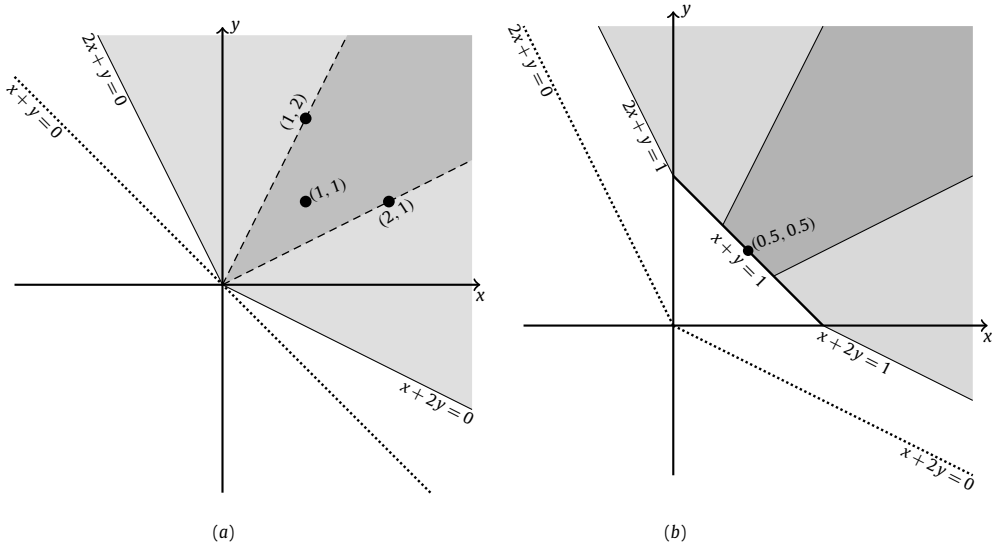


Fig. 1. (a) The visual representation of Example 1 when $\Lambda = \{(2,1),(1,2),(1,1)\}$. If the element γ is in (i) the darkly shaded region; (ii) the first quadrant; (iii) all the shaded region; and (iv) the positive half space of $x + y = 0$ then γ will dominate $\mathbf{0}$ under relation (i) $\succcurlyeq_{\Lambda}^{C}$; (ii) $\succcurlyeq_{\Lambda}^{F}$ and $\succcurlyeq_{\Lambda}^{I,F}$ (these two are equal in this example); (iii) $\succcurlyeq_{\Lambda}^{I}$; and (iv) $\succcurlyeq_{\Lambda}^{mm}$, respectively. **(b)** $\omega_{\Lambda}^{\star}$ equals $(0.5, 0.5)$ $\Lambda^{\geq}$ is the union of the two shaded regions, $\mathrm{SIF}(\Lambda)$ is the part of $\Lambda^{\geq}$ that is strictly within the first quadrant (so not including the axes), $\mathrm{SF}(\Lambda)$ is the part of the line segment $x + y = 1$ strictly within the first quadrant, and $\mathrm{SI}(\Lambda)$ is the darkly shaded region, which is the intersection of $\Lambda^{\geq}$ with $co(\Lambda)$, which is the dark region in the left-hand figure.

set, $\Lambda^{>}$, is shown in Fig. 1(b) as the convex open space above and to the right of the dotted lines, which is also shown as the union of the shaded regions (excluding its boundary) in Fig. 1(a). In Fig. 1(a), the dotted line ($x + y = 0$) is a feasible hyperplane since it could be associated with a feasible point, such as $(0.5, 0.5)$.

2.1. Consistency based relation

One natural preference relation, $\succcurlyeq^{\mathrm{C}}_{\Lambda}$, which has been explored, for example, in [47], is given as follows: the test vector α is *consistency-based preferred* to β ($\alpha \succcurlyeq^{\mathrm{C}}_{\Lambda} \beta$) if and only if $w \cdot \alpha \geq w \cdot \beta$ for all feasible $w \in \Lambda^{>}$. This means that dominance of α over β is consistent with the fact that for all $i \in I$, a_i has dominated b_i.

Proposition 1 below states two other alternative ways to determine if $\alpha \succcurlyeq^{\mathrm{C}}_{\Lambda} \beta$ (just consider $\gamma = \alpha - \beta$). We recall the definition of $\Lambda^{>}$ and define for any finite $\Lambda \subseteq \mathbb{R}^n$, the following three sets:

- $\Lambda^{>} = \{w \in \mathbb{R}^n : \forall \lambda \in \Lambda, w \cdot \lambda > 0\}$;
- $\Lambda^{\geq} = \{w \in \mathbb{R}^n : \forall \lambda \in \Lambda, w \cdot \lambda \geq 1\}$ ($\Lambda^{\geq}$ is the union of the two shaded regions in Fig. 1(b)); and
- $\Lambda^{*} = \{w \in \mathbb{R}^n : \forall \lambda \in \Lambda, w \cdot \lambda \geq 0\}$ (the closed convex space surrounded by dotted lines in Fig. 1(b));
- $co(\Lambda)$, the convex cone generated by Λ, is the smallest convex cone containing Λ (this is the darkly shaded region in Fig. 1(a)); i.e., the set of all vectors in $\mathbb{R}^n$ that can be written as $\sum_{\lambda \in \Lambda} r_\lambda \lambda$, where r_λ are arbitrary non-negative reals. Elements of $co(\Lambda)$ are said to be *positive linear combinations of elements of* Λ.

Proposition 1. *Consider any finite $\Lambda \subseteq \mathbb{R}^n$ that is consistent (i.e., $\Lambda^{>} \neq \emptyset$) and consider any $\gamma \in \mathbb{R}^n$. Then, the following conditions are equivalent. Thus, any of these are equivalent to $\gamma \succcurlyeq^{\mathrm{C}}_{\Lambda} \vec{\mathbf{0}}$.*

(i) *for all $w \in \Lambda^{>}$, $w \cdot \gamma \geq 0$.*
(ii) *for all $w \in \Lambda^{\geq}$, $w \cdot \gamma \geq 0$.*
(iii) $\gamma \in co(\Lambda)$.

2.2. Maximum margin preference relation

The maximum margin preference is based on the principal idea in conventional SVM for the hard margin case (see Section 2.3 of [5] and Section 2 of [6])[2]; This involves picking a unique element w in the feasible set, to generate the preference relation $\geq_w$ given by $\alpha \geq_w \beta \iff w \cdot \alpha \geq w \cdot \beta$ (leading to a stronger ordering than $\succcurlyeq^{\mathrm{C}}_{\Lambda}$).

As mentioned above, w is said to be feasible if $w \cdot \lambda > 0$ for all λ in the set of preference inputs Λ. We can also consider degrees of *feasibility* or *satisfaction*: one might consider $w \cdot \lambda$ as a measure of the degree to which w satisfies λ. However, for our purposes, vector w is equivalent with any scalar multiple of w, such as $2w$, so we want the degree of satisfaction not to be affected by scalar multiplication of w. For this reason, we define the degree $DegSat(w; \lambda)$ that w satisfies λ to be $\frac{w}{\|w\|} \cdot \lambda$. The *margin* $\mathrm{marg}_{\Lambda}(w)$ is then defined as the minimal degree of satisfaction over all elements of Λ, i.e., $\mathrm{marg}_{\Lambda}(w) = \min_{\lambda \in \Lambda} DegSat(w; \lambda) = \min_{\lambda \in \Lambda} \frac{w \cdot \lambda}{\|w\|}$. Note that $\mathrm{marg}_{\Lambda}(w) > 0$ if and only if w is feasible, i.e., $w \in \Lambda^{>}$, and that for any real $r > 0$, $\mathrm{marg}_{\Lambda}(rw) = \mathrm{marg}_{\Lambda}(w)$.

It is natural to choose w that maximises the margin, since it is, in a certain sense, maximally consistent with the preference inputs, i.e., it maximises the degree of satisfaction. The margin $\mathrm{marg}_{\Lambda}(w)$ is equal to the perpendicular distance between the hyperplane H_w and the closest element of Λ to H_w. In simple terms, maximising the margin means choosing a feasible hyperplane that is as far as possible from Λ. The hyperplane that produces the maximum margin is equal to the hyperplane H_w where w uniquely has the minimum (Euclidean) norm in $\Lambda^{\geq}$, as stated in Theorem 2. We denote the unique element of $\Lambda^{\geq}$ with the minimum norm by $\omega^{\star}_{\Lambda}$. In Fig. 1(b), $(0.5, 0.5)$ has uniquely minimal norm in $\Lambda^{\geq}$, so $\omega^{\star}_{\Lambda} = (0.5, 0.5)$, and thus, the associated hyperplane for that point, $x + y = 0$ in Fig. 1(a), has the maximum margin. We use $\|w\|$ as the notation for Euclidean norm in this paper.

Theorem 2. *Let $\Lambda \subseteq \mathbb{R}^n$ be a finite consistent set of preference inputs, so that $\Lambda^{>}$ is non-empty. Then the following all hold.*

(i) *$\Lambda^{\geq}$ is non-empty;*
(ii) *there exists a unique element $\omega^{\star}_{\Lambda}$ in $\Lambda^{\geq}$ with minimum norm;*
(iii) *w maximises marg_{Λ} within $\Lambda^{>}$ if and only if w is a strictly positive scalar multiple of $\omega^{\star}_{\Lambda}$, i.e., there exists $r \in \mathbb{R}$ with $r > 0$ such that $w = r\omega^{\star}_{\Lambda}$.*

More general versions of this result that allow additional linear restrictions on the feasible set $\Lambda^{>}$ are given in [48,49].

[2] It also corresponds to a hard margin version of Ranking SVM [9]; in particular, when the slack variables are omitted or set to zero; this also corresponds, roughly speaking, to tending the penalising constant C in the objective function (Equation 12 of [9]) to infinity..

Theorem 2 allows the following definition of the max-margin preference relation $\succcurlyeq_\Lambda^{mm}$, and also implies that $\alpha \succcurlyeq_\Lambda^{mm} \beta$ if and only if $w \cdot \alpha \geq w \cdot \beta$ for any w maximising the margin, i.e., with maximum degree of satisfaction of the preference inputs.

Definition 2 ($\succcurlyeq_\Lambda^{mm}$). For finite consistent set of preference inputs $\Lambda \subseteq \mathbb{R}^n$ we define relation $\succcurlyeq_\Lambda^{mm}$ by, for $\alpha, \beta \in \mathbb{R}^n$, α is *max-margin-preferred* to β with respect to Λ (i.e., $\alpha \succcurlyeq_\Lambda^{mm} \beta$) if and only if $\alpha \cdot \omega_\Lambda^\star \geq \beta \cdot \omega_\Lambda^\star$, where $\omega_\Lambda^\star$ has uniquely minimal norm in $\Lambda^\geq$.

The relation $\succcurlyeq_\Lambda^{mm}$ is a total pre-order, since it is transitive and for any $\alpha, \beta \in \mathbb{R}^n$ we have $\alpha \succcurlyeq_\Lambda^{mm} \beta$ or $\beta \succcurlyeq_\Lambda^{mm} \alpha$ (or both). Considering Λ as in Example 1, $\omega_\Lambda^\star = (1/2, 1/2)$, and $\{(1/2, 1/2)\}^* = \{(x, y) : x + y \geq 0\}$; so, for all γ in the positive half-space of $x + y = 0$, $\gamma \succcurlyeq_\Lambda^{mm} \vec{\mathbf{0}}$.

2.3. Overall view of rescaling methods

Here we give an overall view of the rescaling approaches developed in the next three sections, Sections 3, 4 and 5. Recall that we are ideally trying to pick an element w of the feasible set $\Lambda^>$, since for any $w \in \Lambda^>$, the associated relation $\geq_w$, given by $\alpha \geq_w \beta \iff \alpha \cdot w \geq \beta \cdot w$, is then consistent with the preference inputs Λ. Note that if we multiply w by a positive constant $r > 0$, this does not change the relation, i.e., the relation $\geq_{rw}$ is equal to the relation $\geq_w$. This implies that we do not lose anything if we focus on the subset $\Lambda^\geq$ of the feasible set $\Lambda^>$ (since they generate the same set of relations $\geq_w$).

As discussed in the previous subsection, the max-margin preference relation (see Definition 2) chooses the element $\omega_\Lambda^\star$ with minimum norm in $\Lambda^\geq$, leading to the associated relation $\succcurlyeq_\Lambda^{mm}$ equalling $\geq_{\omega_\Lambda^\star}$. The effect of different kinds of rescaling leads us to consider different elements of $\Lambda^\geq$, each with an associated subset S of $\Lambda^\geq$; we call S, the *set of scenarios*. We then consider preferences that hold for each element of S; thus, the associated preference relation $\succcurlyeq_S$ is given by $\alpha \succcurlyeq_S \beta$ if and only if for all $w \in S$, $\alpha \cdot w \geq \beta \cdot w$. Equivalently, relation $\succcurlyeq_S$ is equal to $\bigcap_{w \in S} \geq_w$. In the next few paragraphs we discuss different choices for the set of scenarios S.

Max-margin case Regarding the maximum margin preference relation from Section 2.2, $\succcurlyeq_\Lambda^{mm}$ involves just a single scenario, $\omega_\Lambda^\star$, the element of $\Lambda^\geq$ with minimum norm, which uniquely maximises the margin. Recall, α is max-margin-preferred to β if and only if $\alpha \geq_{\omega_\Lambda^\star} \beta$.

Consistency-based For the consistency-based relation $\succcurlyeq_\Lambda^{\mathrm{C}}$ in Section 2.1 we have the set of scenarios S as the whole of $\Lambda^\geq$. Thus, $\succcurlyeq_\Lambda^{\mathrm{C}}$ is the same as $\succcurlyeq_{\Lambda^\geq}$.

Rescaling input preferences Rescaling a preference input, $\lambda \in \Lambda$, means replacing λ by $t_\lambda \lambda$, where t_λ is a strictly positive real, as discussed in Section 3. So, when we rescale the preference inputs, we obtain a new version $\Lambda_{\mathbf{t}}$ of Λ, which has a corresponding element $\omega_{\Lambda_{\mathbf{t}}}^\star$ maximising the margin in the transformed problem. We let $S = \mathrm{SI}(\Lambda)$, the set of all such $\omega_{\Lambda_{\mathbf{t}}}^\star$, leading to relation $\succcurlyeq_\Lambda^{\mathrm{I}}$ that is invariant to the rescaling of inputs. We have that $\alpha \succcurlyeq_\Lambda^{\mathrm{I}} \beta$ if and only if α is max-margin-preferred to β over all rescalings of preference inputs. $\mathrm{SI}(\Lambda)$ is the darkly shaded region in Fig. 1(b).

Rescaling of features A rescaling of the features' domains (as considered in Section 4) amounts to a scaling of each co-ordinate, and thus is associated with a vector τ in $\mathbb{R}^n$ with strictly positive values; e.g., doubling the value of the first feature and leaving the rest unchanged, leads to the rescaling vector τ being $(2, 1, 1, \ldots, 1)$. This transformation affects both the preference inputs Λ, and arbitrary feature vectors, such as α and β. We can then consider the max-margin relation in the transformed space. The *features-scaling-invariant* preference relation $\succcurlyeq_\Lambda^{\mathrm{F}}$ is given by $\alpha \succcurlyeq_\Lambda^{\mathrm{F}} \beta$ if and only if α is max-margin-preferred to β over all rescalings τ of features. The set S of scenarios in this case is equal to the set of what we call the rescale-optimal elements of $\Lambda^\geq$, which are those elements that have minimum rescaled norm for some rescaling τ. The set of rescale-optimal elements is $\mathrm{SF}(\Lambda)$, which equals the part of the line segment $x + y = 1$ strictly within the first quadrant in Fig. 1(b).

Rescaling of both inputs and features In Section 5 we consider both rescaling of the preference inputs and of the features' domains. $\alpha \succcurlyeq_\Lambda^{\mathrm{I,F}} \beta$ if and only if for all rescalings of the features and the preference inputs, α is max-margin preferred to β. This is if and only if $\alpha \geq_w \beta$ for all w in the associated set of scenarios $\mathrm{SIF}(\Lambda)$, where the latter set consists of all ways of transforming $\omega_\Lambda^\star$ by inputs and features rescaling. In Fig. 1(b), $\mathrm{SIF}(\Lambda)$ is the part of the shaded regions that is strictly within the first quadrant.

For each of these sets S of scenarios, we have that $\alpha \not\succcurlyeq_S \beta$ if and only if there exists some $w \in S$, such that $(\beta - \alpha) \cdot w > 0$. As shown in Proposition 1 in Section 2.1, for the case of the consistency-based relation $\succcurlyeq_\Lambda^{\mathrm{C}}$, the simple structure of $S = \Lambda^\geq$ leads to a simple formulation for testing $\alpha \succcurlyeq_S \beta$ that can be solved using linear programming.

However, for the sets of scenarios, $\mathrm{SI}(\Lambda)$, $\mathrm{SF}(\Lambda)$ and $\mathrm{SIF}(\Lambda)$, computation of the associated dominance relations $\succcurlyeq_\Lambda^{\mathrm{I}}$, $\succcurlyeq_\Lambda^{\mathrm{F}}$ and $\succcurlyeq_\Lambda^{\mathrm{I,F}}$ is not straight-forward, because of the more complex definitions. Most of the technical work in Sections 3, 4 and 5

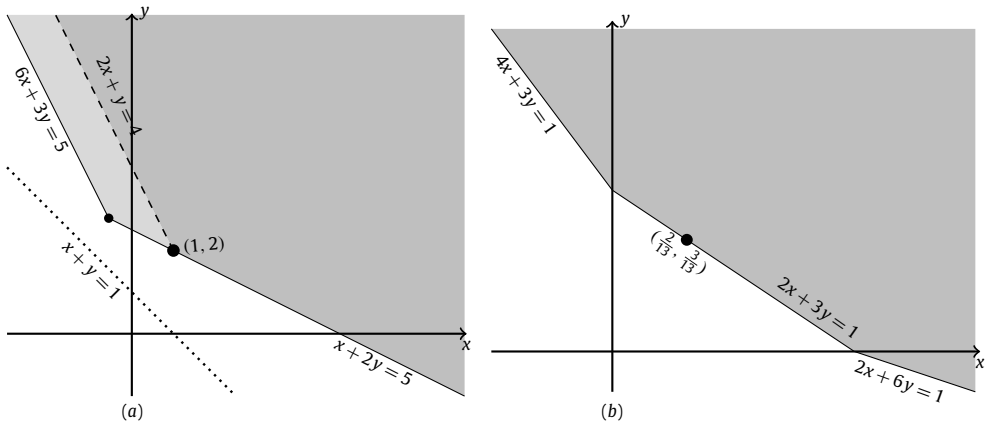


Fig. 2. Some rescalings of inputs and features on $\Lambda = \{(2,1),(1,2),(1,1)\}$ are illustrated. **(a)** $\Lambda_{\mathbf{t}}^{\geq}$ is the union of the two shaded regions when inputs are rescaled by $\mathbf{t} = (3/5, 1/5, 1)$. The darkly shaded region indicates $\Lambda_{\mathbf{t}'}^{\geq}$ where $\mathbf{t}' = (1/4, 1/5, 1/3)$. We have $\omega^{\star}_{\Lambda_{\mathbf{t}}} = \omega^{\star}_{\Lambda_{\mathbf{t}'}} = (1,2)$. The darkly shaded region is also equal to $\Lambda^* + \{(1,2)\}$ (see Example 2 in Section 3.1). **(b)** The shaded region indicates $(\Lambda \odot \tau)^{\geq}$ when features are rescaled by $\tau = (2,3)$. Here, $\omega^{\star}_{\Lambda\odot\tau} = (2/13, 3/13)$ (see Example 3 in Section 4).

is concerned with giving characterisations of the associated sets of scenarios (as constraints involving additional variables) that enable computation of the dominance relations.

3. Rescaling of preference inputs

As discussed in the introduction, a plausible robustness requirement is that a preference relation should not depend on how the preference inputs are scaled. If the user tells us that they prefer α to β, we might expect that this would mean that they would also prefer 0.5α over 0.5β, since we are assuming a linear model. However, if we add this preference, corresponding to $0.5(\alpha - \beta)$, to the preference input set, we may well obtain a different preference relation for the max-margin preference relation; similarly if we replace the original preference $\alpha - \beta$ with this rescaled version $0.5(\alpha - \beta)$. In this section we define and give a characterisation of a preference relation $\succcurlyeq^{\mathrm{I}}_{\Lambda}$ that is invariant to rescaling of the preference inputs Λ.

3.1. Defining inputs-rescaling-invariant relation

Consider the effect of rescaling the preference inputs Λ by $\mathbf{t} \in \mathbb{R}^m_+$ (where $\mathbb{R}^m_+$ is the set of strictly positive reals in m-dimensional), with each preference input being multiplied by a strictly positive scalar, so that the rescaled preference input set is defined as $\Lambda_{\mathbf{t}} = \{\mathbf{t}(i)\lambda_i : i \in I\}$. We then have $(\Lambda_{\mathbf{t}})^{\geq} = \Lambda_{\mathbf{t}}^{\geq} = \{w \in \mathbb{R}^n : \forall i \in I, w \cdot (\mathbf{t}(i)\lambda_i) \geq 1\}$. We will write $\mathbf{t}(i)$ as $\mathbf{t}_i$ for brevity. Note that $(\Lambda_{\mathbf{t}})^{>} = \Lambda^{>}$ for any $\mathbf{t} \in \mathbb{R}^m_+$, since $w \cdot \mathbf{t}_i\lambda_i > 0 \iff w \cdot \lambda_i > 0$, so if Λ is consistent then so is $\Lambda_{\mathbf{t}}$ for every $\mathbf{t} \in \mathbb{R}^m_+$.

Let us say that *α is max-margin-preferred to β under rescaling* $\mathbf{t}$ if $\alpha \succcurlyeq^{mm}_{\Lambda_t} \beta$. Now, it can easily happen that α is preferred to β under one rescaling, but not under another.

Example 2. Consider $\mathbf{t} = (3/5, 1/5, 1)$ rescaling Λ in Example 1. Then, $\Lambda_{\mathbf{t}}$ equals $\{(6/5, 3/5), (1/5, 2/5), (1,1)\}$. In Fig. 2(a), $\Lambda_{\mathbf{t}}^{\geq}$ is the whole shaded region, and it can be seen that $\omega^{\star}_{\Lambda_{\mathbf{t}}} = (1,2)$ which means the hyperplane with the maximum margin for $\Lambda_{\mathbf{t}}$ is $x + 2y = 0$ (instead of $x + y = 0$). Then, $(2, -1.5) \succcurlyeq^{mm}_{\Lambda} (0,0)$ because $(2,-1.5)\cdot(1,1) = 0.5 > 0$, whereas $(2,-1.5) \not\succcurlyeq^{mm}_{\Lambda_t} (0,0)$ because $(2,-1.5)\cdot(1,2) = -1 < 0$.

However, it seems natural to assume that if the user prefers a_i over b_i then he will also prefer $\mathbf{t}_i a_i$ over $\mathbf{t}_i b_i$ for any $\mathbf{t}_i \in \mathbb{R}_+$. Also, for test vectors α and β, if $\alpha \succcurlyeq^{mm}_{\Lambda} \beta$ then, for any positive real r, we have $r\alpha \succcurlyeq^{mm}_{\Lambda} r\beta$; since the resultant preferences are invariant to such rescaling, it seems reasonable that the same would hold for the input preferences.

We therefore consider a more robust relation, which is invariant to the scaling of the preference inputs, with α being *inputs-scaling-invariant* preferred to β only if it is max-margin preferred for all rescalings $\mathbf{t} \in \mathbb{R}^m_+$ of the preference inputs.

Definition 3 ($\succcurlyeq^{\mathrm{I}}_{\Lambda}$). For finite consistent set of preference inputs $\Lambda \subseteq \mathbb{R}^n$ we define relation $\succcurlyeq^{\mathrm{I}}_{\Lambda}$ by, for $\alpha, \beta \in \mathbb{R}^n$, $\alpha \succcurlyeq^{\mathrm{I}}_{\Lambda} \beta$ if and only if α is max-margin-preferred to β over all rescalings of preference inputs, i.e., if for all $\mathbf{t} \in \mathbb{R}^m_+$, $\alpha \succcurlyeq^{mm}_{\Lambda_{\mathbf{t}}} \beta$.

So far, we have assumed that each component $\mathbf{t}_i$ of $\mathbf{t}$ can be any strictly positive scalar. However, in Proposition 3 below, we will show that if each $\mathbf{t}_i$ is restricted to be in $(0,1]$, the result for relation $\succcurlyeq^{\mathrm{I}}_{\Lambda}$ will not change. This is not

surprising, since, e.g., doubling each component of $\mathbf{t}$ will not change the relation $\succcurlyeq^{mm}_{\Lambda_{\mathbf{t}}}$. This simplification will be helpful in the computation of the $\succcurlyeq^{\mathrm{I}}_{\Lambda}$ relation.

Proposition 3. *Consider any finite consistent set of preference inputs $\Lambda \subseteq \mathbb{R}^n$ and any $\alpha, \beta \in \mathbb{R}^n$. Then, $\alpha \succcurlyeq^{\mathrm{I}}_{\Lambda} \beta$ if and only if for all $\mathbf{t} \in (0,1]^m$, $\alpha \succcurlyeq^{mm}_{\Lambda_{\mathbf{t}}} \beta$.*

We define $\mathrm{SI}(\Lambda)$ to be the set consisting solely of $\omega^{\star}_{\Lambda_{\mathbf{t}}}$ for all scalings $\mathbf{t} \in (0,1]^m$.

Definition 4 ($\mathrm{SI}(\Lambda)$)*.* For finite consistent set of preference inputs $\Lambda \subseteq \mathbb{R}^n$, let $\mathrm{SI}(\Lambda) = \{\omega^{\star}_{\Lambda_{\mathbf{t}}} : \mathbf{t} \in (0,1]^m\}$.

Thus, $u \in \mathrm{SI}(\Lambda)$ if and only if there exists $\mathbf{t} \in (0,1]^m$ such that u has minimal norm in $\Lambda^{\geq}_{\mathbf{t}}$. Proposition 3, along with Definition 2, immediately implies the following result, expressing the preference relation $\succcurlyeq^{\mathrm{I}}_{\Lambda}$ in terms of the set $\mathrm{SI}(\Lambda)$.

Proposition 4. *For finite consistent set of preference inputs $\Lambda \subseteq \mathbb{R}^n$ and $\alpha, \beta \in \mathbb{R}^n$, we have $\alpha \succcurlyeq^{\mathrm{I}}_{\Lambda} \beta \iff$ for all $w \in \mathrm{SI}(\Lambda)$, $\alpha \cdot w \geq \beta \cdot w$.*

For example, it can be shown that $\mathrm{SI}(\Lambda)$ in Fig. 1(b) is the darkly shaded region, which is the intersection of the shaded region $\Lambda^{\geq}$ in Fig. 1(b) with $co(\Lambda)$, which is the dark region in the left-hand figure (see Theorem 7 below). Then, it can be seen that $(\mathrm{SI}(\Lambda))^*$ is all the shaded region in Fig. 1(a). This implies that $\gamma \succcurlyeq^{\mathrm{I}}_{\Lambda} \vec{\mathbf{0}}$ if and only if γ is in any of the shaded region in Fig. 1(a). Also, $\alpha \succcurlyeq^{\mathrm{I}}_{\Lambda} \beta \iff \alpha - \beta \succcurlyeq^{\mathrm{I}}_{\Lambda} \vec{\mathbf{0}}$.

3.2. Characterisation of $\mathrm{SI}(\Lambda)$

Here, we mathematically characterise $\mathrm{SI}(\Lambda)$; this will lead to a computational method for the $\succcurlyeq^{\mathrm{I}}_{\Lambda}$ relation. Proposition 5 below implies that $\mathrm{SI}(\Lambda) \subseteq \Lambda^{\geq}$ and for every element u in $\mathrm{SI}(\Lambda)$, vector u has minimum norm in $\Lambda^* + \{u\} = \{w + u : w \in \Lambda^*\}$ (which equals $\{w' \in \mathbb{R}^n : \forall \lambda \in \Lambda, w' \cdot \lambda \geq u \cdot \lambda\}$).

In the running example, assume $u = (1,2)$. Then, $\Lambda^* + \{u\} = \{(x,y) : 2x + y \geq 4, x + 2y \geq 5, x + y \geq 3\}$ which is the darkly shaded region in Fig. 2(a) (place the origin on u and then draw Λ^*).

Proposition 5. *Consider a finite consistent set of preference inputs $\Lambda \subseteq \mathbb{R}^n$ and any $u \in \mathbb{R}^n$. Then, $u \in \mathrm{SI}(\Lambda)$ if and only if $u \in \Lambda^{\geq}$ and u has minimum norm in $\Lambda^* + \{u\}$. Thus, in particular, $\mathrm{SI}(\Lambda) \subseteq \Lambda^{\geq}$.*

We know that $u = (1,2) \in \mathrm{SI}(\Lambda)$ because it has minimum norm in $\Lambda_{\mathbf{t}}$ for $\mathbf{t} = (3/5, 1/5, 1)$. We can easily see that $u \in \Lambda^{\geq}$ and has minimum norm in $\Lambda^* + \{u\}$. Now, let v be any point between two black circles in Fig. 2(a). Then, v does not have minimal norm in $\Lambda^* + \{v\}$; in fact, $(1,2)$ minimises norm in $\Lambda^* + \{v\}$. We will see that $v \notin \mathrm{SI}(\Lambda)$.

We will prove (in Proposition 6) that $co(\Lambda)$ is precisely the set of elements $u \in \mathbb{R}^n$ such that u has minimum norm in $\Lambda^* + \{u\}$. Together with Proposition 5, this will imply Theorem 7 below, which characterises SI.

Proposition 6. *Consider any finite consistent set of preference inputs $\Lambda \subseteq \mathbb{R}^n$ and any $u \in \mathbb{R}^n$. Then, u has minimum norm in $\Lambda^* + \{u\}$ if and only if $u \in co(\Lambda)$.*

Propositions 5 and 6 immediately imply the following theorem.

Theorem 7. *Consider any finite consistent set of preference inputs $\Lambda \subseteq \mathbb{R}^n$. Then, $\mathrm{SI}(\Lambda) = co(\Lambda) \cap \Lambda^{\geq}$.*

Proof. u is in $\mathrm{SI}(\Lambda)$ if and only if by Proposition 5, $u \in \Lambda^{\geq}$ and u has minimum norm in $\Lambda^* + \{u\}$, which, from Proposition 6, holds if and only if $u \in \Lambda^{\geq}$ and $u \in co(\Lambda)$. □

Theorem 7 shows that $\mathrm{SI}(\Lambda)$ in Fig. 1(b) is the darkly shaded region, which is the intersection of the shaded region $\Lambda^{\geq}$ in Fig. 1(b) with $co(\Lambda)$, which is the dark region in Fig. 1(a).

The following result leads immediately to an algorithm to determine, for arbitrary $\alpha, \beta \in \mathbb{R}^n$ if $\alpha \succcurlyeq^{\mathrm{I}}_{\Lambda} \beta$, using a linear programming solver.

Corollary 8. *For finite consistent set of preference inputs $\Lambda \subseteq \mathbb{R}^n$, let $\lambda_i \in \Lambda$ be the i^{th} element of Λ where $i \in I = \{1, \ldots, |\Lambda|\}$. Consider any $u \in \mathbb{R}^n$. Then, u is in $\mathrm{SI}(\Lambda)$ if and only if for all $i \in I$, $u \cdot \lambda_i \geq 1$ and there exist non-negative reals r_i for each $i \in I$ such that $u = \sum_{i \in I} r_i \lambda_i$.*

Proof. The result follows easily from Theorem 7 and the definition of $co(\Lambda)$ and $\Lambda^{\geq}$. □

Proposition 4 implies that for $\alpha, \beta \in \mathbb{R}^n$, $\alpha \not\succcurlyeq^{\mathrm{I}}_{\Lambda} \beta$ if and only if there exists $u \in \mathrm{SI}(\Lambda)$ such that $\alpha \cdot u < \beta \cdot u$. Using Corollary 8 we therefore have the following result which leads immediately to a computational procedure for the preference relation $\succcurlyeq^{\mathrm{I}}_{\Lambda}$.

Proposition 9. *Let $\Lambda \subseteq \mathbb{R}^n$ be a finite consistent set of preference inputs, and let $\alpha, \beta \in \mathbb{R}^n$. Then $\alpha \not\succcurlyeq^{\mathrm{I}}_{\Lambda} \beta$ if and only if there exists $u \in \mathbb{R}^n$ such that the following three conditions all hold:*

(i) $u \cdot (\beta - \alpha) > 0$;
(ii) *for all $i \in I$, $u \cdot \lambda_i \geq 1$; and*
(iii) *there exist non-negative reals r_i for each $i \in I$ such that $u = \sum_{i \in I} r_i \lambda_i$.*

4. Rescaling of features

As discussed in the introduction, an important, and potentially problematic, pre-processing step in SVM methods is rescaling of the domain of each feature. In this section we define a preference relation $\succcurlyeq^{\mathrm{F}}_{\Lambda}$ (based on preference inputs Λ) that is invariant to the relative scalings of the feature domains.

Normalization of features is a necessary phase in any SVM-based method. This task often involves translations and rescalings on the domain of each feature. It is evident that the maximum margin relation is unaffected by translation of feature space; i.e., for all $\delta \in \mathbb{R}^n$, $\alpha + \delta \succcurlyeq^{mm}_{\Lambda} \beta + \delta$ iff $(\alpha + \delta) \cdot \omega^{\star}_{\Lambda} \geq (\beta + \delta) \cdot \omega^{\star}_{\Lambda}$ if and only if $\alpha \succcurlyeq^{mm}_{\Lambda} \beta$. Therefore, in this section we only consider the effect of rescaling of feature spaces.

The effect of rescaling of features on a conventional binary SVM classifier is also discussed in a separate study by the authors [50]. In that context, a data point is called *strongly positive* (respectively negative) if it is positively (resp. negatively) classified for all choices of feature scaling. Otherwise, the instance is considered *neutral* because it is differently classified for different scalings of features.

Let $\mathbb{R}^n_+$ be the set of strictly positive vectors in $\mathbb{R}^n$, i.e., vectors with every component strictly positive. Let *features rescaling* $\tau \in \mathbb{R}^n_+$ be a vector of strictly positive reals, with the jth component $\tau(j)$ being the scale factor for the jth feature. The effect of the rescaling on a vector $\lambda \in \mathbb{R}^n$ is given by pointwise multiplication, $\lambda \odot \tau$, defined by, for all $j = 1, \ldots, n$, $(\lambda \odot \tau)(j) = \lambda(j)\tau(j)$. Operation $\odot$ is commutative, associative and distributes over addition of vectors. An important property is that for any $u, v, w \in \mathbb{R}^n$ $(u \odot v) \cdot w = v \cdot (u \odot w)$, since they are both equal to $\sum_{j=1}^{n} u(j)v(j)w(j)$. For $\tau \in \mathbb{R}^n_+$, we define τ^{-1} to be the element of $\mathbb{R}^n_+$ given by $\tau^{-1}(j) = 1/\tau(j)$ for all $j \in \{1, \ldots, n\}$. The rescaling vector changes the preference inputs Λ, turning it into $\Lambda \odot \tau = \{\lambda \odot \tau : \lambda \in \Lambda\}$. Let $\omega^{\star}_{\Lambda \odot \tau}$ be the element with minimum norm in $(\Lambda \odot \tau)^{\geq}$, where $(\Lambda \odot \tau)^{\geq} = \{w \in \mathbb{R}^n : \forall \lambda \in \Lambda, w \cdot (\lambda \odot \tau) \geq 1\}$.

Example 3. Consider $\tau = (2, 3)$ rescaling features of Λ in Example 1. Then, $\Lambda \odot \tau$ will be $\{(4, 3), (2, 6), (2, 3)\}$. The shaded region in Fig. 2(b) shows $(\Lambda \odot \tau)^{\geq}$, and it can be seen that $\omega^{\star}_{\Lambda \odot \tau} = (2/13, 3/13)$. Then, $(2, -1) \odot \tau \not\succcurlyeq^{mm}_{\Lambda \odot \tau} (0, 0) \odot \tau$ because $(2, -1) \odot (2, 3) \cdot (2/13, 3/13) = -1/13 < 0$, whereas $(2, -1) \succcurlyeq^{mm}_{\Lambda} (0, 0)$ because $(2, -1) \cdot (1/2, 1/2) = 1/2 > 0$.

Like rescaling of inputs, we see that α might be preferred to β under one rescaling of features, but not under another. However, the choice of how the features are scaled relative to each other can involve somewhat arbitrary choices. It is therefore natural to consider a more cautious preference relation, *features-scaling-invariant* preference relation, given by α being preferred to β for all rescalings $\tau \in \mathbb{R}^n_+$.

Definition 5 *($\succcurlyeq^{\mathrm{F}}_{\Lambda}$).* For finite consistent set of preference inputs $\Lambda \subseteq \mathbb{R}^n$ we define relation $\succcurlyeq^{\mathrm{F}}_{\Lambda}$ by, for $\alpha, \beta \in \mathbb{R}^n$, $\alpha \succcurlyeq^{\mathrm{F}}_{\Lambda} \beta$ if and only if α is max-margin-preferred to β over all rescalings of features, i.e., if for all $\tau \in \mathbb{R}^n_+$, we have $\alpha \odot \tau \succcurlyeq^{mm}_{\Lambda \odot \tau} \beta \odot \tau$.

We define the set of vectors $\mathrm{SF}(\Lambda)$ as follows.

Definition 6 *($\mathrm{SF}(\Lambda)$).* For finite consistent set of preference inputs $\Lambda \subseteq \mathbb{R}^n$, define $\mathrm{SF}(\Lambda)$ to be $\{\omega^{\star}_{\Lambda \odot \tau} \odot \tau : \tau \in \mathbb{R}^n_+\}$.

We then have the following simple relationship between $\mathrm{SF}(\Lambda)$ and the preference relation $\succcurlyeq^{\mathrm{F}}_{\Lambda}$.

Proposition 10. *For finite consistent set of preference inputs $\Lambda \subseteq \mathbb{R}^n$, and any $\alpha, \beta \in \mathbb{R}^n$, we have $\alpha \succcurlyeq^{\mathrm{F}}_{\Lambda} \beta \iff$ for all $w \in \mathrm{SF}(\Lambda)$, $\alpha \cdot w \geq \beta \cdot w$.*

Proof. We have that $\alpha \succcurlyeq^{\mathrm{F}}_{\Lambda} \beta \iff$ for all $\tau \in \mathbb{R}^n_+$, $\alpha \odot \tau \succcurlyeq^{mm}_{\Lambda \odot \tau} \beta \odot \tau$. Now, $\alpha \odot \tau \succcurlyeq^{mm}_{\Lambda \odot \tau} \beta \odot \tau$ if and only if $(\alpha \odot \tau) \cdot \omega^{\star}_{\Lambda \odot \tau} \geq (\beta \odot \tau) \cdot \omega^{\star}_{\Lambda \odot \tau}$, i.e., $\alpha \cdot (\omega^{\star}_{\Lambda \odot \tau} \odot \tau) \geq \beta \cdot (\omega^{\star}_{\Lambda \odot \tau} \odot \tau)$. Thus, $\alpha \succcurlyeq^{\mathrm{F}}_{\Lambda} \beta \iff$ for all $\tau \in \mathbb{R}^n_+$, $\alpha \cdot (\omega^{\star}_{\Lambda \odot \tau} \odot \tau) \geq \beta \cdot (\omega^{\star}_{\Lambda \odot \tau} \odot \tau)$, which is if and only if for all $w \in \mathrm{SF}(\Lambda)$, $\alpha \cdot w \geq \beta \cdot w$. □

4.1. Rescale optimality

We define an important notion, rescale optimality, for understanding the set $\mathrm{SF}(\Lambda)$, and hence the features-scaling-invariant preference relation $\succcurlyeq^{\mathrm{F}}_{\Lambda}$. We will see below, in Proposition 11, that $\mathrm{SF}(\Lambda)$ is equal to the set of rescale optimal elements of $\Lambda^{\geq}$. Because some of the formal concepts and results do not require exactly the form of $\Lambda^{\geq}$, we express them in terms of a more general subset G of $\mathbb{R}^n$.

Definition 7 *(Rescale-optimal).* For $G \subseteq \mathbb{R}^n$, and $u \in G$, let us say that u *is rescale-optimal in* G if there exists (strictly positive) $\tau \in \mathbb{R}^n_+$ with $\|\tau \odot w\| \geq \|\tau \odot u\|$ for all $w \in G$.

It can be seen intuitively that elements of the (open) line segment between $(1,0)$ and $(0,1)$ in Fig. 1(b) is the set of rescale-optimal elements in $\Lambda^{\geq}$; if $\tau(1) > \tau(2)$ (i.e., with the ratio $\tau(1)/\tau(2)$ being increased) then $\omega^{\star}_{\Lambda\odot\tau} \odot \tau$ moves from $(1/2, 1/2)$ towards $(1,0)$. Similarly, increasing the ratio $\tau(2)/\tau(1)$ from 1 moves $\omega^{\star}_{\Lambda\odot\tau} \odot \tau$ from $(1/2, 1/2)$ towards $(0,1)$.

We will show in Proposition 11 below that $\mathrm{SF}(\Lambda)$ is equal to the set of rescale-optimal elements in $\Lambda^{\geq}$. For instance, $\omega^{\star}_{\Lambda\odot\tau} \odot \tau$ in Example 3 (where $\tau(2)/\tau(1) = 3/2$), which is in $\mathrm{SF}(\Lambda)$ by the definition, is equal to $(2/13, 3/13) \odot (2,3) = (4/13, 9/13)$, that is between $(1/2, 1/2)$ and $(0,1)$. If $\mathrm{SF}(\Lambda)$ equals the line segment between $(1,0)$ and $(0,1)$, it can be seen that $(\mathrm{SF}(\Lambda))^*$ is the first quadrant in Fig. 1(a). This implies that in Fig. 1(a), $\gamma \succcurlyeq^{\mathrm{F}}_{\Lambda} \vec{\mathbf{0}}$ if and only if γ is in the first quadrant.

Proposition 11. *Consider any finite consistent set of preference inputs $\Lambda \subseteq \mathbb{R}^n$. Then, $\mathrm{SF}(\Lambda)$ is equal to the set of all rescale-optimal elements of $\Lambda^{\geq}$. Thus, for $\alpha, \beta \in \mathbb{R}^n$, $\alpha \succcurlyeq^{\mathrm{F}}_{\Lambda} \beta$ if and only if $w \cdot (\alpha - \beta) \geq 0$ for every rescale-optimal element w in $\Lambda^{\geq}$.*

Proposition 11 implies, in particular, that $\omega^{\star}_{\Lambda}$ is rescale-optimal in $\Lambda^{\geq}$.

4.2. Pointwise undominated

Let G be some subset of $\mathbb{R}^n$ (where we will be applying the results to the case of $G = \Lambda^{\geq}$). For $u \in G$, if there exists $v \in G$ such that for all j, $v(j)$ is between $u(j)$ and 0 then it is easy to see that u cannot be rescale-optimal element in G. This is the idea behind being pointwise undominated, which is reminiscent of being Pareto undominated, and is a necessary condition for being rescale-optimal. The notion of pointwise dominance leads to a characterisation of when there is a unique rescale-optimal element, see Theorem 13 in Section 4.3, which corresponds to the case in which rescaling of features makes no difference.

Definition 8 *(pointwise dominance).* For $u, v \in \mathbb{R}^n$, v pointwise dominates u if $u \neq v$ and for all $j \in \{1, \ldots, n\}$, either

(i) $0 \leq v(j) \leq u(j)$, or
(ii) $0 \geq v(j) \geq u(j)$.

For $u \in G \subseteq \mathbb{R}^n$, we say that u is *pointwise undominated* in G if there exists no $v \in G$ that pointwise dominates u.

In Fig. 1(b), all elements on the part of *closed* line segment $x + y = 1$ within the first quadrant (i.e., including points on the axes) are pointwise undominated in $\Lambda^{\geq}$. The definition easily implies that being rescale-optimal implies being pointwise undominated (but not the converse).

Proposition 12. *Let $G \subseteq \mathbb{R}^n$. If u is rescale-optimal in G then u is pointwise undominated in G. Thus, if u is pointwise dominated in G then u is not rescale-optimal in G.*

Proof. Suppose that u is not pointwise undominated in G, so that there exists $v \in G$ that pointwise dominates u. Then, for every $j \in \{1, \ldots, n\}$, $|v(j)| \leq |u(j)|$, and for some $k \in \{1, \ldots, n\}$, $|v(k)| < |u(k)|$, which implies that for every $\tau \in \mathbb{R}^n_+$, $\|v \odot \tau\| < \|u \odot \tau\|$, and hence, u is not rescale-optimal in G. □

Proposition 12 states that being pointwise undominated is a necessary condition for being rescale-optimal. However, by having a look at our running example we will see that this not a sufficient condition. The intersection points of $x + y = 1$ with the axes (i.e., $(1,0)$ and $(0,1)$) are pointwise undominated but not rescale-optimal in $\Lambda^{\geq}$. To see this, suppose that for example $(1,0)$ were rescale-optimal in $\Lambda^{\geq}$; i.e., there exists $\tau \in \mathbb{R}^2_+$ such that for all $v \in \Lambda^{\geq} - \{(1,0)\}$, $\|(1,0) \odot \tau\| < \|v \odot \tau\|$. In particular, there exists $\tau \in \mathbb{R}^2_+$ such that for all $\epsilon \in (0,1]$, $\|(1,0) \odot \tau\| < \|(1-\epsilon, \epsilon) \odot \tau\|$, i.e., $\tau^2(x) < (1-\epsilon)^2\tau^2(x) + \epsilon^2\tau^2(y)$. Letting $r = \tau(x)/\tau(y)$, we obtain, there exists $r \in \mathbb{R}_+$ such that for all $\epsilon \in (0,1]$, $r^2 < (1-\epsilon)^2 r^2 + \epsilon^2$, and thus, $r^2 < \epsilon^2/(1-(1-\epsilon)^2) = \epsilon/(2-\epsilon)$. Now, for any $r \in \mathbb{R}_+$ there exists sufficiently small $\epsilon > 0$ such that $\epsilon/(2-\epsilon) < r^2$, proving that $(1,0)$ is not rescale-optimal in $\Lambda^{\geq}$ by contradiction. We can use a similar argument to show that $(0,1)$ is not rescale-optimal in $\Lambda^{\geq}$ as well. We will investigate this further in Section 4.4, leading to a computational procedure for rescale-optimality. First, in Section 4.3, we characterise the situations when rescaling of features makes no difference, in which case $\succcurlyeq^{\mathrm{F}}_{\Lambda}$ is the same as $\succcurlyeq^{mm}_{\Lambda}$.

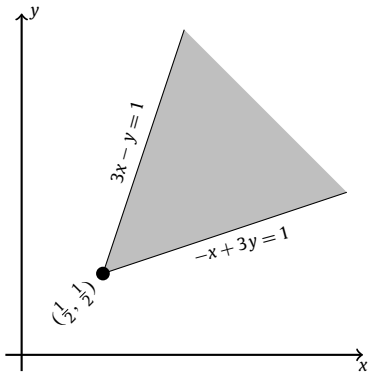


Fig. 3. The visual representation of Example 4 when $\Lambda = \{(-1,3),(3,-1)\}$. $\Lambda^{\geq}$ is the shaded region with a single extremal point at $(1/2, 1/2)$. In this case, $\succcurlyeq^{\mathrm{F}}_{\Lambda}$ is equal to $\succcurlyeq^{mm}_{\Lambda}$.

4.3. Determining invariance to rescaling of features

Example 4 below illustrates that allowing rescaling of features can sometimes make no difference in maximum margin relation.

Example 4. Consider Λ be $\{(-1,3),(3,-1)\}$ so that $\Lambda^{\geq} = \{(x,y) : -x+3y \geq 1, 3x-y \geq 1\}$ which is the shaded region in Fig. 3. Here, $\Lambda^{\geq}$ has a single extremal point at $(1/2, 1/2)$. Since $(1/2, 1/2)$ is the element with minimal norm in $\Lambda^{\geq}$, $\omega^{\star}_{\Lambda} = (1/2, 1/2)$, and so $(1/2, 1/2)$ is rescale-optimal in $\Lambda^{\geq}$. Also, all other points in $\Lambda^{\geq}$ are pointwise dominated by $(1/2, 1/2)$; thus, by Proposition 12, they are not rescale-optimal. Consequently, the only element of $\Lambda^{\geq}$ that is rescale-optimal is $(1/2, 1/2)$.

Note that if there exists a unique rescale-optimal element in $\Lambda^{\geq}$, then this element must be $\omega^{\star}_{\Lambda}$, since the latter is rescale-optimal by Proposition 11. This immediately implies that $\succcurlyeq^{\mathrm{F}}_{\Lambda}$ is then equal to $\succcurlyeq^{mm}_{\Lambda}$. Therefore this is the situation in which rescaling of the features has no effect on the preference relation.

Theorem 13 below states that u is the only rescale-optimal element in convex closed G if and only if u pointwise dominates every other element of G.

Theorem 13. *Let G be a convex and closed subset of $\mathbb{R}^n$, and let u be an element of G. Then the following conditions are equivalent.*

(i) *u is uniquely rescale-optimal in G, i.e., u is the unique element of G that is rescale-optimal;*
(ii) *for all $v \in G$, for all $j \in \{1,\ldots,n\}$, $|v(j)| \geq |u(j)|$;*
(iii) *u pointwise dominates every element in $G - \{u\}$.*

Consider Λ as it is in Example 4. Then, the three conditions hold for $u = (1/2, 1/2)$. The equivalence between (i) and (ii) is proved using Lemmas 42 and 43, and the equivalence between (ii) and (iii) follows using Lemma 15.

Corollary 14. *Let $\Lambda \subseteq \mathbb{R}^n$ be a finite consistent set of preference inputs. Choose an arbitrary element $y \in \Lambda^{\geq}$. Using y we will generate an element $y^* \in \mathbb{R}^n$. For each $j \in \{1,\ldots,n\}$: If $y(j) = 0$ then let $y^*(j) = 0$. If $y(j) > 0$ then let $y^*(j) = \inf\{w(j) : w \in \Lambda^{\geq}, w(j) \geq 0\}$. If $y(j) < 0$ then let $y^*(j) = \sup\{w(j) : w \in \Lambda^{\geq}, w(j) \leq 0\}$. If $y^* \in \Lambda^{\geq}$ then y^* is uniquely rescale-optimal in $\Lambda^{\geq}$. Also, there exists a uniquely rescale-optimal element in $\Lambda^{\geq}$ if and only if $y^* \in \Lambda^{\geq}$.*

Consider Λ as in Example 4. Choose $y = (1,1)$ which is in $\Lambda^{\geq}$. Since $y(1) = 1 > 0$, $y^*(1) = \inf\{w(1) : w \in \Lambda^{\geq}, w(1) \geq 0\} = \frac{1}{2}$. Similarly, $y(2) = 1$ implies that $y^*(2) = \frac{1}{2}$. Because $(\frac{1}{2}, \frac{1}{2}) \in \Lambda^{\geq}$, it is uniquely rescale-optimal in $\Lambda^{\geq}$. If you consider Λ as in Example 1, then $y = (1,1)$ will be updated to $y^* = (0,0)$, and because $(0,0) \notin \Lambda^{\geq}$, there does not exist a uniquely rescale-optimal element in $\Lambda^{\geq}$.

To prove the corollary we use the following lemma.

Lemma 15. *Let G be a convex subset of $\mathbb{R}^n$, and let j be any element of $\{1,\ldots,n\}$. Then either (i) there exists $w \in G$ such that $w(j) = 0$; or (ii) for all $w \in G$, $w(j) > 0$; or (iii) for all $w \in G$, $w(j) < 0$.*

Proof of Corollary 14. Consider any $j \in \{1,\ldots,n\}$. Lemma 15 implies that if $y(j) > 0$ then for all $w \in \Lambda^{\geq}$, $0 < y^*(j) \leq w(j)$; and if $y(j) < 0$ then for all $w \in \Lambda^{\geq}$, $0 > y^*(j) \geq w(j)$, so for all $w \in \Lambda^{\geq}$, either $0 \leq y^*(j) \leq w(j)$ or $0 \geq y^*(j) \geq w(j)$. Theorem 13 implies that if $y^* \in \Lambda^{\geq}$ then y^* is uniquely rescale-optimal in $\Lambda^{\geq}$.

Conversely, suppose that there exists a uniquely rescale-optimal element u in $\Lambda^{\geq}$; we will prove that $y^* \in \Lambda^{\geq}$. Consider arbitrary $j \in \{1, \ldots, n\}$. The fact that $\Lambda^{\geq}$ is a polyhedron implies that there exists some $w \in \Lambda^{\geq}$ with $w(j) = y^*(j)$. If $y^*(j) > 0$ then we know by Lemma 15 that $u(j) \geq y^*(j)$. But Theorem 13 implies that $u(j) \leq w(j) = y^*(j)$, and thus $y^*(j) = u(j)$. Similarly, if $y^*(j) < 0$ then $y^*(j) \geq u(j) \geq w(j) = y^*(j)$ and so, $y^*(j) = u(j)$. If $y^*(j) = 0$ then $w(j) = 0$, so $u(j) = 0$, also using Theorem 13. We have shown that $y^* = u$, so $y^* \in \Lambda^{\geq}$. □

Corollary 14 leads immediately to an algorithm for determining if $\Lambda^{\geq}$ has a uniquely rescale-optimal element, and finding it, if it exists. This is the situation in which rescaling of the features makes no difference. The algorithm involves at most $n+1$ runs of a linear programming solver, and thus determining and finding a uniquely rescale-optimal element u can be performed in polynomial time. If it succeeds in finding such a u then the induced preferences can be efficiently tested using: $\alpha \succcurlyeq^{\mathrm{F}}_{\Lambda} \beta$ if and only if $u \cdot (\alpha - \beta) \geq 0$.

4.4. Characterising rescale-optimality

As we have shown, being pointwise undominated is a necessary but not a sufficient condition for being rescale-optimal. In this section we define a stronger version of pointwise undominated called *zm-pointwise undominated*, where 'zm' stands for *zeros-modified* (the essential difference being in the treatment of j such that $u(j) = 0$). We show that this is a necessary condition as well, and is in fact also a sufficient condition for being rescale-optimal (for polyhedra). According to the following definition, while the points $(1, 0)$ and $(0, 1)$ in Fig. 1(b) are pointwise undominated, they are not zm-pointwise undominated.

Definition 9 *(zm-pointwise Dominance).* For $u, v \in \mathbb{R}^n$, let $N_u = \{j \in \{1, \ldots, n\} : u(j) \neq 0\}$. v zm-pointwise dominates u if there exists $k \in N_u$ such that $u(k) \neq v(k)$ and for all $j \in N_u$, either (i) $0 \leq v(j) \leq u(j)$, or (ii) $0 \geq v(j) \geq u(j)$.

For $u \in G \subseteq \mathbb{R}^n$, we say that u is *zm-pointwise undominated* in G if there exists no $v \in G$ that zm-pointwise dominates u.

Clearly, if every component of u is non-zero, then $N_u = \{1, \ldots, n\}$, and so for any vector $v \in \mathbb{R}^n$ we have that v zm-pointwise dominates u if and only if v pointwise dominates u. Proposition 16 below gives a characterisation of (i) rescale-optimal, and (ii) zm-pointwise undominated. Together, these immediately imply part (iii), that being zm-pointwise undominated is a necessary condition for being rescale-optimal.

Proposition 16. *Let u be an element of convex $G \subseteq \mathbb{R}^n$. Then:*

(i) *u is rescale-optimal in G if and only if there exists $\tau \in \mathbb{R}^n_+$ such that for all $v \in G$, $(\tau \odot \tau \odot u) \cdot (v - u) \geq 0$ (i.e., $(\tau \odot u) \cdot (\tau \odot (v - u)) \geq 0$).*
(ii) *u is zm-pointwise undominated in G if and only if for all $v \in G$, there exists $\tau \in \mathbb{R}^n_+$ such that $(\tau \odot \tau \odot u) \cdot (v - u) \geq 0$.*
(iii) *If u is rescale-optimal in G then u is zm-pointwise undominated in G.*

We say that $u, v \in \mathbb{R}^n$ *agree on signs* if, for each component j, $u(j)$ and $v(j)$ have equal sign.

Definition 10 *(Agreeing on Signs).* For $u, v \in \mathbb{R}^n$, *u and v agree on signs* if for all $j = 1, \ldots, n$,

(i) $u(j) = 0 \iff v(j) = 0$;
(ii) $u(j) > 0 \iff v(j) > 0$; and thus also:
(iii) $u(j) < 0 \iff v(j) < 0$.

For example, $(1, 0)$ and $(1, 1)$ do not agree on signs but for $\varepsilon > 0$, $(1, \varepsilon)$ and $(1, 1)$ agree on signs. Clearly, if u and v agree on signs then $u \cdot v > 0$, unless they're both the zero vector. The following is the key theorem of this section to characterise rescale-optimality by making use of Proposition 16(i). This characterisation is the basis of the computational procedure for the features-scaling-invariant preference relation $\succcurlyeq^{\mathrm{F}}_{\Lambda}$ developed in Section 4.6.

Theorem 17. *Consider any u in convex $G \subseteq \mathbb{R}^n$. If $u = \vec{\mathbf{0}}$ then it is the unique rescale-optimal element of G. Otherwise, u is rescale-optimal in G if and only if there exists $\mu \in \mathbb{R}^n$ agreeing on signs with u such that $\mu \cdot u = 1$ and for all $w \in G$, $\mu \cdot w \geq 1$.*

To illustrate this result, consider u be any element of the part of the open line segment $x + y = 1$ within the first quadrant in Fig. 1(b); i.e., $u = (\delta, 1 - \delta)$ for some $\delta \in (0, 1)$. Then for $\mu = (1, 1)$, μ and u agree on signs and $u \cdot \mu = 1$, and for all $w \in \Lambda^{\geq}$, $\mu \cdot w \geq 1$. Therefore, u is rescale-optimal in $\Lambda^{\geq}$.

Proof. It is clear that if $u = \vec{\mathbf{0}}$ then for all $\tau \in \mathbb{R}^n_+$ and for all $w \in G - \{u\}$, $\|u \odot \tau\| = 0 < \|w \odot \tau\|$, which means that u is the unique rescale-optimal element of G. Now, suppose $u \neq \vec{\mathbf{0}}$.

$\Rightarrow$: First, let us assume that u is rescale-optimal in G. Then, by Proposition 16(i), there exists $\tau \in \mathbb{R}^n_+$ such that for all $w \in G$, $(\tau \odot \tau \odot u) \cdot (w - u) \geq 0$. Let $\mu' = \tau \odot \tau \odot u$. Because, $u \neq \vec{\mathbf{0}}$, $\mu' \cdot u = \|\tau \odot u\|^2 > 0$. Then, we let $\mu = \frac{\mu'}{\mu' \cdot u}$ which leads to $\mu \cdot u = 1$. In addition, μ and μ' agree on signs because $\mu' \cdot u > 0$, and μ' and u agree on signs, by the definition of μ', and hence μ and u agree on signs. Consider any $w \in G$. We know that $(\tau \odot \tau \odot u) \cdot (w - u) \geq 0$, i.e., $\mu' \cdot (w - u) \geq 0$, which implies that $\mu \cdot (w - u) \geq 0$, and therefore, $\mu \cdot w \geq 1 = \mu \cdot u$.

$\Leftarrow$: For the converse, we assume that there exists $\mu \in \mathbb{R}^n$ agreeing on signs with u such that $\mu \cdot u = 1$ and for all $w \in G$, $\mu \cdot w \geq 1$, and thus, $\mu \cdot (w - u) \geq 0$. Define $\tau \in \mathbb{R}^n$ by $\tau(j) = 1$ if $u(j) = 0$, and $\tau(j) = \sqrt{\frac{\mu(j)}{u(j)}}$ whenever $u(j) \neq 0$, which is well-defined, because $\mu(j)/u(j) > 0$ whenever $u(j) \neq 0$, using the fact that μ and u agree on signs. That fact also means that $u(j) = 0$ implies that $\mu(j) = 0$. We then have that for all $j \in \{1, \ldots, n\}$, $\tau(j)^2 u(j) = \mu(j)$, which implies that $\tau \odot \tau \odot u = \mu$. Therefore, for all $w \in G$, $(\tau \odot \tau \odot u) \cdot (w - u) \geq 0$, which implies that u is rescale-optimal in G, by Proposition 16(i). □

4.5. Equivalence of rescale-optimal with zm-pointwise undominated

It turns out that being zm-pointwise undominated is equivalent to being rescale-optimal, for a polyhedron, and therefore, in particular, for $\Lambda^{\geq}$; see Theorem 18 below. The rather complex proof of this result can be found in the appendix.

Theorem 18. *Let u be an element of polyhedron $G \subseteq \mathbb{R}^n$. Then, u is rescale-optimal in G if and only if u is zm-pointwise undominated in G.*

4.6. Expressing rescale-optimality in terms of positive linear combinations

Here we extend the characterisation of rescale-optimality given in Theorem 17, leading to a computational method for testing rescale-optimality, and thus to a method for testing if $\alpha \succcurlyeq^{\mathrm{F}}_{\Lambda} \beta$, for $\alpha, \beta \in \mathbb{R}^n$, i.e., preference with respect to the features-scaling-invariant preference relation. Theorem 17 implies that non-zero u is rescale-optimal in convex set G if and only if there exists a vector μ that agrees on signs with u with $\mu \cdot w \geq \mu \cdot u$ for all $w \in G$. Theorem 21 below shows that μ is a positive linear combination of certain vectors when G is a polyhedron, which thus includes the case when $G = \Lambda^{\geq}$. This leads to a characterisation in Theorem 23 of $\mathrm{SF}(\Lambda)$, giving a computation procedure for the preference relation $\succcurlyeq^{\mathrm{F}}_{\Lambda}$, summed up in Proposition 24.

We can write any polyhedron as $G_I = \{w \in \mathbb{R}^n : \forall i \in I, w \cdot \lambda_i \geq a_i\}$, for finite I, and with each $\lambda_i \in \mathbb{R}^n$ and $a_i \in \mathbb{R}$. We also consider $G_{J_u} = \{w \in \mathbb{R}^n : \forall i \in J_u, w \cdot \lambda_i \geq a_i\}$, where $J_u = \{i \in I : \lambda_i \cdot u = a_i\}$. Clearly, for all $u \in G_I$, $G_I \subseteq G_{J_u}$.

For example, consider $a_1 = a_2 = 1$, $u = (1, 0)$, $v = (1/2, 1/2)$ and $y = (1, 1)$, with the vectors λ_i for $i \in I = \{1, 2, 3\}$ being as in Example 1 and Fig. 1. Then, $G_{J_u} = \{w \in \mathbb{R}^2 : w \cdot (1, 1) \geq 1, w \cdot (1, 2) \geq 1\}$; $G_{J_v} = \{w \in \mathbb{R}^2 : w \cdot (1, 1) \geq 1\}$; and $G_{J_y} = \mathbb{R}^2$ because $J_y = \emptyset$.

The following pair of lemmas are used in the proof, with the first one following very easily from the definition.

Lemma 19. *$G_{J_u} + \{-u\}$ is equal to $\{\lambda_i : i \in J_u\}^*$.*

Lemma 20. *Consider a polyhedron G_I and non-zero $u \in G_I$. Then u is rescale-optimal in G_I if and only if u is rescale-optimal in G_{J_u}.*

Theorem 21. *Let G be a polyhedron, which we write as $G_I = \{w \in \mathbb{R}^n : \forall i \in I, w \cdot \lambda_i \geq a_i\}$, for finite I, and with each $\lambda_i \in \mathbb{R}^n$ and $a_i \in \mathbb{R}$. Consider any non-zero vector u in G_I. Then, u is rescale-optimal in G_I if and only if there exists $\mu \in \mathbb{R}^n$ that agrees on signs with u such that $\mu \cdot u = 1$ and $\mu \in co(\{\lambda_i : i \in J_u\})$.*

Recall the example illustrating Theorem 17 where $u = (\delta, 1 - \delta)$ for some (arbitrary) $\delta \in (0, 1)$. We can see that the set $\{\lambda_i : i \in J_u\}$ equals $\{(1, 1)\}$. So, choosing $\mu = (1, 1)$ gives that u is rescale-optimal in $\Lambda^{\geq}$ because u and μ agree on signs, $u \cdot \mu = 1$, and $\mu \in co(\{(1, 1)\})$.

Note that this theorem implies that if non-zero u is rescale-optimal in G_I then J_u is non-empty, since $\vec{\mathbf{0}}$ is the only positive linear combination of the empty set, and $\mu \neq \vec{\mathbf{0}}$.

Proof. First consider $\mu \in \mathbb{R}^n$ such that $\mu \cdot u = 1$. Then it can be seen that $\{w : w \cdot \mu \geq 1\} + \{-u\} = (\{\mu\})^*$. Also, $G_{J_u} \subseteq \{w : w \cdot \mu \geq 1\}$ if and only if $G_{J_u} + \{-u\} \subseteq (\{\mu\})^* \iff \{\lambda_i : i \in J_u\}^* \subseteq (\{\mu\})^*$, using Lemma 19, which is if and only if, $\mu \in co(\{\lambda_i : i \in J_u\})$ (because of a standard result—see Lemma 33).

By Lemma 20, u is rescale-optimal in G_I if and only if u is rescale-optimal in G_{J_u}, which, by Theorem 17, is if and only if there exists $\mu \in \mathbb{R}^n$ agreeing on signs with u such that $\mu \cdot u = 1$ and $G_{J_u} \subseteq \{w : w \cdot \mu \geq 1\}$, i.e., $\mu \in co(\{\lambda_i : i \in J_u\})$, by the earlier argument. □

We have the following corollary (using the same notation), which shows that testing if u is rescale-optimal in G_I can be performed in polynomial time: by first checking that $u \in G_I$ (i.e., for all $i \in I$, $u \cdot \lambda_i \geq a_i$), and then testing if a set of inequalities has a solution, using a linear programming solver.

Corollary 22. *Let u be a non-zero element of $\mathbb{R}^n$. Then, u is rescale-optimal in G_I if and only if $u \in G_I$ and there exists non-negative reals r_i for each $i \in J_u$, and vector $\tau \in \mathbb{R}^n$ with for all $j \in \{1,\ldots,n\}$, $\tau(j) \geq 1$, and $\tau(j)u(j) = \sum_{i\in J_u} r_i\lambda_i(j)$.*

Proof. First suppose that u is rescale-optimal in G_I. Then $u \in G_I$, and, by Theorem 21, there exists $\mu \in \mathbb{R}^n$ that agrees on signs with u such that $\mu \cdot u = 1$ and there exist non-negative $r'_i \in \mathbb{R}$ such that $\mu = \sum_{i\in J_u} r'_i\lambda_i$. For all $j \in \{1,\ldots,n\}$ such that $u(j) \neq 0$, define $t_j = \mu(j)/u(j)$, which is greater than zero, because μ and u agree on signs, and let t be the minimum of these values. Define τ by $\tau(j) = 1$ if $u(j) = 0$ and otherwise, define $\tau(j) = t_j/t$. Then for all $j \in \{1,\ldots,n\}$, $\tau(j) \geq 1$, and $\tau(j)u(j) = \mu(j)/t = \sum_{i\in J_u}(r'_i/t)\lambda_i(j)$.

Conversely, suppose that $u \in G_I$ and there exists non-negative reals r_i for each $i \in J_u$ and vector $\tau \in \mathbb{R}^n$ with for all $j \in \{1,\ldots,n\}$, $\tau(j) \geq 1$, and $\tau(j)u(j) = \sum_{i\in J_u} r_i\lambda_i(j)$. Define $\mu \in \mathbb{R}^n$ to be $\frac{\tau\odot u}{(\tau\odot u)\cdot u}$. Then $\mu \cdot u = 1$, and μ agrees on signs with u, and is a positive linear combination of $\{\lambda_i : i \in J_u\}$. Theorem 21 then can be applied to give the result. □

Theorem 21 implies the following, which leads to a computational method for checking dominance with respect to $\succcurlyeq^{\mathrm{F}}_{\Lambda}$.

Theorem 23. *Consider finite consistent set of preference inputs $\Lambda \subseteq \mathbb{R}^n$, and any $u \in \mathbb{R}^n$. Define $\Theta_u = \{\lambda \in \Lambda : \lambda \cdot u = 1\}$. Then, u is in $\mathrm{SF}(\Lambda)$ if and only if $u \in \Lambda^{\geq}$ and there exists $\mu \in \mathbb{R}^n$ such that μ agrees on signs with u, and $\mu \in co(\Theta_u)$. Also, u is in $\mathrm{SF}(\Lambda)$ if and only if $u \in \Lambda^{\geq}$ and there exists $\mu \in \mathbb{R}^n$ and some subset Δ of Θ_u such that $|\Delta| \leq n+1$, and $\mu \in co(\Delta)$, and μ agrees on signs with u.*

Proof. Proposition 11 implies that $\mathrm{SF}(\Lambda)$ equals the set of all rescale-optimal elements of $\Lambda^{\geq}$. Hence, Theorem 21 implies that $u \in \mathrm{SF}(\Lambda)$ if and only if $u \in \Lambda^{\geq}$ and there exists $\mu \in \mathbb{R}^n$ that agrees on signs with u such that $\mu \in co(\Theta_u)$ and $\mu \cdot u = 1$. First, we will show that the condition $\mu \cdot u = 1$ can be omitted.

Suppose first that $u \in \Lambda^{\geq}$ and there exists $\mu \in \mathbb{R}^n$ that agrees on signs with u such that $\mu \in co(\Theta_u)$. Now, u is a non-zero vector, since $u \in \Lambda^{\geq}$. Since μ agrees on signs with u, we have$\mu \cdot u > 0$. Define $\mu' = \frac{\mu}{\mu\cdot u}$. Then $\mu' \in co(\Lambda)$, $\mu' \cdot u = 1$, and μ' and u agree on signs. We can then apply Theorem 21 to give $u \in \mathrm{SF}(\Lambda)$. The converse follows immediately from the same theorem.

The last part follows from Carathéodory's Theorem (see e.g., 3.1.2 in [51]) which states that for any $w \in \mathbb{R}^n$ and any $S \subseteq \mathbb{R}^n$, if $w \in co(S)$ then there exists $S' \subseteq S$ with $|S'| \leq n+1$ such that $w \in co(S')$. □

Proposition 10 implies that for $\alpha, \beta \in \mathbb{R}^n$, $\alpha \not\succcurlyeq^{\mathrm{F}}_{\Lambda} \beta$ if and only if there exists $u \in \mathrm{SF}(\Lambda)$ such that $\alpha \cdot u < \beta \cdot u$. Theorem 23 then implies the following characterisation, leading to a computational procedure for the $\succcurlyeq^{\mathrm{F}}_{\Lambda}$ relation (see Section 7.2 below).

Proposition 24. *Let $\Lambda \subseteq \mathbb{R}^n$ be a finite consistent set of preference inputs, and let $\alpha, \beta \in \mathbb{R}^n$. We have $\alpha \not\succcurlyeq^{\mathrm{F}}_{\Lambda} \beta$ if and only if there exists $u \in \mathbb{R}^n$ and $\mu \in \mathbb{R}^n$ such that the following four conditions all hold:*

(i) $u \cdot (\beta - \alpha) > 0$;
(ii) *for all $i \in I$, $u \cdot \lambda_i \geq 1$ (i.e., $u \in \Lambda^{\geq}$);*
(iii) *for all $j = \{1,\ldots,n\}$, $u(j) = 0 \iff \mu(j) = 0$, and $u(j) > 0 \iff \mu(j) > 0$ (i.e., agreeing on signs); and*
(iv) *there exists some subset Δ of $\Theta_u = \{\lambda \in \Lambda : \lambda \cdot u = 1\}$ such that $|\Delta| \leq n+1$, and there exist non-negative reals r_i for each $i \in I$ such that $\mu = \sum_{i\in I} r_i\lambda_i$ where $r_i = 0$ if $\lambda_i \notin \Delta$ (so, $\mu \in co(\Delta)$).*

Condition (iv) holds if and only if there exist non-negative reals r_i for each $i \in I$ such that $\sum_{i\in I}(r_i \neq 0) \leq n+1$ (i.e., $|\{i \in I : r_i \neq 0\}| \leq n+1$), and $\mu = \sum_{i\in I} r_i\lambda_i$, and for all $i \in I$, either $u \cdot \lambda_i = 1$ or $r_i = 0$.

5. Simultaneous rescaling of features and inputs

Having defined the preference relations $\succcurlyeq^{\mathrm{I}}_{\Lambda}$ and $\succcurlyeq^{\mathrm{F}}_{\Lambda}$, based, respectively, on rescaling of preference inputs and features, it is also natural to consider both kinds of rescaling simultaneously. In this section, we define and characterise a preference relation based on allowing both the rescaling of features and of preference inputs.

Definition 11 *($\mathrm{SIF}(\Lambda)$ and $\succcurlyeq^{\mathrm{I,F}}_{\Lambda}$).* For finite consistent set of preference inputs $\Lambda \subseteq \mathbb{R}^n$, we define the set $\mathrm{SIF}(\Lambda)$ by $w \in \mathrm{SIF}(\Lambda)$ if there exists $\mathbf{t} \in (0,1]^m$ such that $w \in \mathrm{SF}(\Lambda_{\mathbf{t}})$; i.e., $\mathrm{SIF}(\Lambda) = \{\omega^{\star}_{\Lambda_{\mathbf{t}}\odot\tau} \odot \tau : \mathbf{t} \in (0,1]^m, \tau \in \mathbb{R}^n_+\}$. We define relation $\succcurlyeq^{\mathrm{I,F}}_{\Lambda}$ by $\alpha \succcurlyeq^{\mathrm{I,F}}_{\Lambda} \beta \iff$ for all $w \in \mathrm{SIF}(\Lambda)$, $w \cdot \alpha \geq w \cdot \beta$.

This definition implies that $\alpha \succcurlyeq^{\mathrm{I,F}}_{\Lambda} \beta$ if and only if for all rescalings of the features and the preference inputs, α is max-margin preferred to β. We have the following characterisation, which leads to a computational method for checking if $\alpha \succcurlyeq^{\mathrm{I,F}}_{\Lambda} \beta$.

Theorem 25. *Let $\Lambda \subseteq \mathbb{R}^n$ be a finite consistent set of preference inputs. Then, $u \in \mathrm{SIF}(\Lambda)$ if and only if $u \in \Lambda^{\geq}$ and there exists $\mu \in \mathbb{R}^n$ that agrees on signs with u such that $\mu \in co(\Lambda)$.*

Consider Λ as in Example 1. For any choice of $\mu \in co(\Lambda)$, we have either $\mu = (0,0)$ or $\mu(1) > 0$ and $\mu(2) > 0$ (see Fig. 1(a)). Because $(0,0) \notin \Lambda^{\geq}$, $u \in \mathrm{SIF}(\Lambda)$ if $u(1) > 0$ and $u(2) > 0$ (to agree on signs with μ); however, $u \neq (0,0)$ because $(0,0) \notin \Lambda^{\geq}$. As a result, $\mathrm{SIF}(\Lambda)$ is the part of $\Lambda^{\geq}$ (i.e., the darkly shaded region) that is *strictly* within the first quadrant in Fig. 1(b). It can be seen that $(\mathrm{SIF}(\Lambda))^*$ is the first quadrant in Fig. 1(a). This implies that for $\gamma \in \mathbb{R}^n$, $\gamma \succcurlyeq_{\Lambda}^{\mathrm{I,F}} \vec{\mathbf{0}}$ if and only if γ is in the first quadrant in Fig. 1(a).

Proof. We first show that $\mathrm{SIF}(\Lambda) \subseteq \Lambda^{\geq}$. We have $\mathrm{SIF}(\Lambda) = \bigcup_{\mathbf{t}} \mathrm{SF}(\Lambda_{\mathbf{t}})$, where the union is over all $\mathbf{t} \in (0,1]^m$. Also, by Proposition 11, $\mathrm{SF}(\Lambda_{\mathbf{t}}) \subseteq \Lambda_{\mathbf{t}}^{\geq}$; it follows easily that $\Lambda_{\mathbf{t}}^{\geq} \subseteq \Lambda^{\geq}$ (see Lemma 39 below), and thus we have $\mathrm{SF}(\Lambda_{\mathbf{t}}) \subseteq \Lambda^{\geq}$. Therefore, $\mathrm{SIF}(\Lambda) \subseteq \Lambda^{\geq}$.

Now suppose that $u \in \mathrm{SIF}(\Lambda)$; as shown above, we then have $u \in \Lambda^{\geq}$. By definition of $\mathrm{SIF}(\Lambda)$ there exists $\mathbf{t} \in (0,1]^m$ such that $u \in \mathrm{SF}(\Lambda_{\mathbf{t}})$. Theorem 23 implies that there exists $\mu \in \mathbb{R}^n$, that agrees on signs with u, such that $\mu \in co(\{\mathbf{t}_i\lambda_i \in \Lambda_{\mathbf{t}} : \mathbf{t}_i\lambda_i \cdot u = 1\})$, and thus, in particular, μ is in $co(\Lambda_{\mathbf{t}})$, which equals $co(\Lambda)$. Hence, there exists $\mu \in \mathbb{R}^n$, that agrees on signs with u, such that $\mu \in co(\Lambda)$.

For the converse, assume that $u \in \Lambda^{\geq}$ and there exists $\mu \in \mathbb{R}^n$, that agrees on signs with u, such that $\mu \in co(\Lambda)$. Let us define $\mathbf{t} \in \mathbb{R}_+^m$ by $\mathbf{t}(i) = \frac{1}{\lambda_i \cdot u}$ for all $i \in \{1, \ldots, m\}$. Because $u \in \Lambda^{\geq}$ we have $\lambda_i \cdot u \geq 1$, and thus, $\mathbf{t}(i) \in (0,1]$. Then, for all i we have $\mathbf{t}_i\lambda_i \cdot u = 1$, which implies that $u \in \Lambda_{\mathbf{t}}^{\geq}$ and also that $\Lambda = \{\lambda_i \in \Lambda : \mathbf{t}_i\lambda_i \cdot u = 1\}$. Now, $co(\{\lambda_i \in \Lambda : \mathbf{t}_i\lambda_i \cdot u = 1\})$ equals $co(\{\mathbf{t}_i\lambda_i \in \Lambda_{\mathbf{t}} : \mathbf{t}_i\lambda_i \cdot u = 1\})$, and hence, $\mu \in co(\{\mathbf{t}_i\lambda_i \in \Lambda_{\mathbf{t}} : \mathbf{t}_i\lambda_i \cdot u = 1\})$. Since, $u \in \Lambda_{\mathbf{t}}^{\geq}$, Theorem 23 implies that $u \in \mathrm{SF}(\Lambda_{\mathbf{t}})$. We therefore have that $u \in \mathrm{SIF}(\Lambda)$. □

This result gives rise a computational procedure for the $\succcurlyeq_{\Lambda}^{\mathrm{I,F}}$ relation; Definition 11 implies that for $\alpha, \beta \in \mathbb{R}^n$, $\alpha \not\succcurlyeq_{\Lambda}^{\mathrm{I,F}} \beta$ if and only if there exists $u \in \mathrm{SIF}(\Lambda)$ such that $\alpha \cdot u < \beta \cdot u$, which leads, by Theorem 25, to following characterisation of $\succcurlyeq_{\Lambda}^{\mathrm{I,F}}$.

Proposition 26. *Let $\Lambda \subseteq \mathbb{R}^n$ be a finite consistent set of preference inputs, and let $\alpha, \beta \in \mathbb{R}^n$. Then, $\alpha \not\succcurlyeq_{\Lambda}^{\mathrm{I,F}} \beta$ if and only if there exists $u, \mu \in \mathbb{R}^n$, such that the following four conditions all hold:*

(i) $u \cdot (\beta - \alpha) > 0$;
(ii) *for all $i \in I$, $u \cdot \lambda_i \geq 1$ (i.e., $u \in \Lambda^{\geq}$);*
(iii) *for all $j = \{1, \ldots, n\}$, $u(j) = 0 \iff \mu(j) = 0$, and $u(j) > 0 \iff \mu(j) > 0$ (i.e., agreeing on signs); and*
(iv) *there exist non-negative reals r_i for each $i \in I$ such that $\mu = \sum_{i \in I} r_i\lambda_i$ (i.e., $\mu \in co(\Lambda)$).*

6. Generating a consistent preference input set

There are a number ways of extending the approach to deal with inconsistent input information, i.e., when $\Lambda^{\geq}$ is empty, where Λ is the (finite) set of preference inputs. One desirable property of such a method is that it should not depend on an arbitrary ordering of the input set Λ. Here, we describe three possible approaches for restoring consistency, which all satisfy this property.

The first approach is iteratively eliminating the elements of Λ that are least consistent with others. Define the function $C : \Lambda \to \mathbb{R}$ such that for every $i \in I$, $C(\lambda_i) = \sum_{j \in I - \{i\}} \lambda_i \cdot \lambda_j$. This function expresses a kind of degree of consistency of the element λ_i with other elements of Λ, where the smaller the value of $C(\lambda_i)$ is, the less consistency there is between λ_i and the other elements of Λ. Then the simple procedure Algorithm 1 can be followed to generate a consistent subset of Λ.

Algorithm 1 Generating Consistent Subset of Λ.

```
function GENERATING CONSISTENT SUBSET(Λ)
    while Λ^≥ = ∅ do                          ▷ Λ is inconsistent
        γ := arg min_{λ∈Λ} C(λ)                ▷ γ minimises C
        Λ := Λ − {γ}                          ▷ Remove γ from Λ
    end while
    return Λ                                  ▷ Λ is now consistent
end function
```

The second method forms a consistent subset of Λ based on the sum $\mu = \sum_{\lambda \in \Lambda} \lambda$ of the preference input vectors: see Proposition 27 below. Unless μ is the zero vector, $\Lambda_\mu = \{\lambda \in \Lambda : \lambda \cdot \mu > 0\}$ is non-empty and consistent. We can therefore define $\omega^\star_{\Lambda_\mu}$ to be the solution of the maximum margin approach for Λ_μ. Then, we return $\Lambda_{\omega^\star_{\Lambda_\mu}} = \{\lambda \in \Lambda : \lambda \cdot \omega^\star_{\Lambda_\mu} > 0\}$ which is again consistent, and we have $\Lambda_\mu \subseteq \Lambda_{\omega^\star_{\Lambda_\mu}} \subseteq \Lambda$.

Proposition 27. *Consider any finite subset Λ of $\mathbb{R}^n$, and define $\mu = \sum_{\lambda \in \Lambda} \lambda$ and define Λ_μ to be $\{\lambda \in \Lambda : \lambda \cdot \mu > 0\}$. Assume that $\mu \neq \vec{\mathbf{0}}$. Then the following hold.*

(i) Λ_μ *is non-empty and consistent, i.e.,* $(\Lambda_\mu)^> \neq \emptyset$.
(ii) *Let* $\omega^\star_{\Lambda_\mu}$ *be the solution of the maximum margin approach for* Λ_μ*, i.e., the minimal norm element in* $(\Lambda_\mu)^\geq$*. Then* $\Lambda_{\omega^\star_{\Lambda_\mu}} = \{\lambda \in \Lambda : \lambda \cdot \omega^\star_{\Lambda_\mu} > 0\}$ *is non-empty and consistent, and* $\Lambda_\mu \subseteq \Lambda_{\omega^\star_{\Lambda_\mu}} \subseteq \Lambda$.

A third approach involves adding m extra real variables (i.e., m dummy features), one for each λ_i (with $i \in I = \{1, \ldots, m\}$) and extend each λ_i to the extra m variables by it having a value ε in the corresponding column, and zeros in the other $m-1$ columns. Here, ε is a strictly positive (typically small) number that relates inversely to the penalty for softening the constraints.

More formally, we say that $u \in \mathbb{R}^{n+m}$ extends $v \in \mathbb{R}^n$ if for each $j = 1, \ldots, n$, $u(j) = v(j)$. For each $i \in I$ we define δ_i as follows: δ_i extends λ_i, and $\delta_i(n+i) = \varepsilon$, and $\delta_i(n+j) = 0$ for $j \in I - \{i\}$. Let Δ, the extended preference inputs set, equal $\{\delta_i : i \in I\}$.

Consider any $w \in \mathbb{R}^n$, and any $u \in \mathbb{R}^{n+m}$ that extends w. Then, for each $i \in I$, $u \cdot \delta_i = w \cdot \lambda_i + \varepsilon u(n+i)$. Thus, $u \cdot \delta_i \geq 1$ if and only if $u(n+i) \geq \frac{1}{\varepsilon}(1 - w \cdot \lambda_i)$. If $w \cdot \lambda_i \geq 1$ then we can satisfy the constraint $u \cdot \delta_i \geq 1$ by setting $u(n+i) = 0$. Otherwise, we can satisfy the constraint by letting $u(n+i) = \frac{1}{\varepsilon}(1 - w \cdot \lambda_i)$. (In fact, since we are interested in minimising the norm, or a rescaled version of the norm, we only need to consider this particular way of extending w to $\mathbb{R}^{n+m}$.)

This implies that any $w \in \mathbb{R}^n$ can be extended to an element of $\Delta^\geq$; so, in particular, the extended input set Δ is always consistent. However, if w is not close to satisfying λ_i, i.e., if $w \cdot \lambda_i$ is a large negative number, then the value of $u(n+i)$, and hence the norm of u, will be large. This shows that vectors $w \in \mathbb{R}^n$ that come close to satisfying the input constraints will be favoured.

The definitions and mathematical machinery for the various preference relations defined above can then proceed as in the previous sections but now working within $\mathbb{R}^{n+m}$. When testing dominance the test vectors α and β are extended with the same value (e.g., 0) for the extra m components.

7. Properties of relations and computation of inferences

In previous sections, we defined a number of preference relations. In Section 7.1 we give some properties, in particular, regarding the relationships between the preference relations. In Section 7.2 we express the computational characterisations, derived in earlier sections, in terms of constraints, which enable simple implementation.

7.1. Properties of the different preference relations

We have considered the following preference relations: the consistency-based relation $\succcurlyeq^{\mathrm{C}}_\Lambda$ (Section 2.1), the relation $\succcurlyeq^{\mathrm{I}}_\Lambda$ based on rescaling preference inputs for the maximum margin preference relation (Section 3), relation $\succcurlyeq^{\mathrm{F}}_\Lambda$ based on rescaling of features (Section 4) and relation $\succcurlyeq^{\mathrm{I,F}}_\Lambda$ based on rescaling both inputs and features (Section 5).

For each of the relations $\succcurlyeq^{\mathrm{C}}_\Lambda, \succcurlyeq^{\mathrm{I}}_\Lambda, \succcurlyeq^{\mathrm{F}}_\Lambda$ and $\succcurlyeq^{\mathrm{I,F}}_\Lambda$, the corresponding *set of scenarios* is defined to be $\mathrm{SC}(\Lambda)$, $\mathrm{SI}(\Lambda)$, $\mathrm{SF}(\Lambda)$ and $\mathrm{SIF}(\Lambda)$, respectively, where $\mathrm{SC}(\Lambda)$ is defined to be $\Lambda^\geq$. For $u \in \mathbb{R}^n$ recall that the total pre-order $\geq_u$ is given by $\alpha \geq_u \beta \iff u \cdot \alpha \geq u \cdot \beta$. Let $\succcurlyeq$ be any of the relations $\succcurlyeq^{\mathrm{C}}_\Lambda, \succcurlyeq^{\mathrm{I}}_\Lambda, \succcurlyeq^{\mathrm{F}}_\Lambda$ and $\succcurlyeq^{\mathrm{I,F}}_\Lambda$ and let S be the corresponding set of scenarios for each relation, so that $\succcurlyeq = \succcurlyeq_S$ using the notation of Section 2.3. We then have that $\succcurlyeq$ is the intersection of relations $\geq_u$ over all $u \in S$: see Section 2.1, and Proposition 4, Proposition 10 and Definition 11.

We also consider the intersection of $\succcurlyeq^{\mathrm{I}}_\Lambda$ and $\succcurlyeq^{\mathrm{F}}_\Lambda$, which we call $\succcurlyeq^{\mathrm{I}\wedge\mathrm{F}}_\Lambda$, so that, for $\alpha, \beta \in \mathbb{R}^n$, $\alpha \succcurlyeq^{\mathrm{I}\wedge\mathrm{F}}_\Lambda \beta$ if and only if $\alpha \succcurlyeq^{\mathrm{I}}_\Lambda \beta$ and $\alpha \succcurlyeq^{\mathrm{F}}_\Lambda \beta$, which is if and only if $\alpha \geq_u \beta$ for all $u \in \mathrm{SI}(\Lambda) \cup \mathrm{SF}(\Lambda)$. Hence, if $S = \mathrm{SI}(\Lambda) \cup \mathrm{SF}(\Lambda)$ then $\succcurlyeq_S$ equals $\succcurlyeq^{\mathrm{I}\wedge\mathrm{F}}_\Lambda = \succcurlyeq^{\mathrm{I}}_\Lambda \cap \succcurlyeq^{\mathrm{F}}_\Lambda$.

These relations, as well as $\succcurlyeq^{mm}_\Lambda$, are all reflexive and transitive, and thus pre-orders (with $\succcurlyeq^{mm}_\Lambda$ being a total pre-order). This is because each relation is equal to an intersection of pre-orders. For similar reasons, the relations are preserved under some simple transformations.

We say that binary relation $\succcurlyeq$ on $\mathbb{R}^n$ is *preserved by translation and uniform positive scaling* if for any $\lambda \in \Lambda$ and for $\alpha, \beta, \gamma \in \mathbb{R}^n$ and $r \in \mathbb{R}_+$, if $\alpha \succcurlyeq \beta$ then $\alpha + \gamma \succcurlyeq \beta + \gamma$ and $r\alpha \succcurlyeq r\beta$.

Proposition 28. *For finite consistent set of preference inputs* $\Lambda \subseteq \mathbb{R}^n$*, we have the following relationships between the sets of scenarios:*

$$\omega^\star_\Lambda \in \mathrm{SI}(\Lambda) \cap \mathrm{SF}(\Lambda) \text{ and } \mathrm{SI}(\Lambda) \cup \mathrm{SF}(\Lambda) \subseteq \mathrm{SIF}(\Lambda) \subseteq \mathrm{SC}(\Lambda) = \Lambda^\geq.$$

Let $\succcurlyeq$ *be any of the relations* $\succcurlyeq^{mm}_\Lambda, \succcurlyeq^{\mathrm{C}}_\Lambda, \succcurlyeq^{\mathrm{I}}_\Lambda, \succcurlyeq^{\mathrm{F}}_\Lambda$ *and* $\succcurlyeq^{\mathrm{I,F}}_\Lambda$*. Then,* $\succcurlyeq$ *is a pre-order preserved by translation and uniform positive scaling, and* $\lambda \succ \vec{\mathbf{0}}$ *for all* $\lambda \in \Lambda$ *(where* $\succ$ *is the strict part of* $\succcurlyeq$*). In addition, these relations are nested in the following ways (see Fig. 4):*

$$\succcurlyeq^{\mathrm{C}}_\Lambda \subseteq \succcurlyeq^{\mathrm{I,F}}_\Lambda \subseteq \succcurlyeq^{\mathrm{I}\wedge\mathrm{F}}_\Lambda = \succcurlyeq^{\mathrm{I}}_\Lambda \cap \succcurlyeq^{\mathrm{F}}_\Lambda, \text{ and } \succcurlyeq^{\mathrm{I}}_\Lambda \cup \succcurlyeq^{\mathrm{F}}_\Lambda \subseteq \succcurlyeq^{mm}_\Lambda.$$

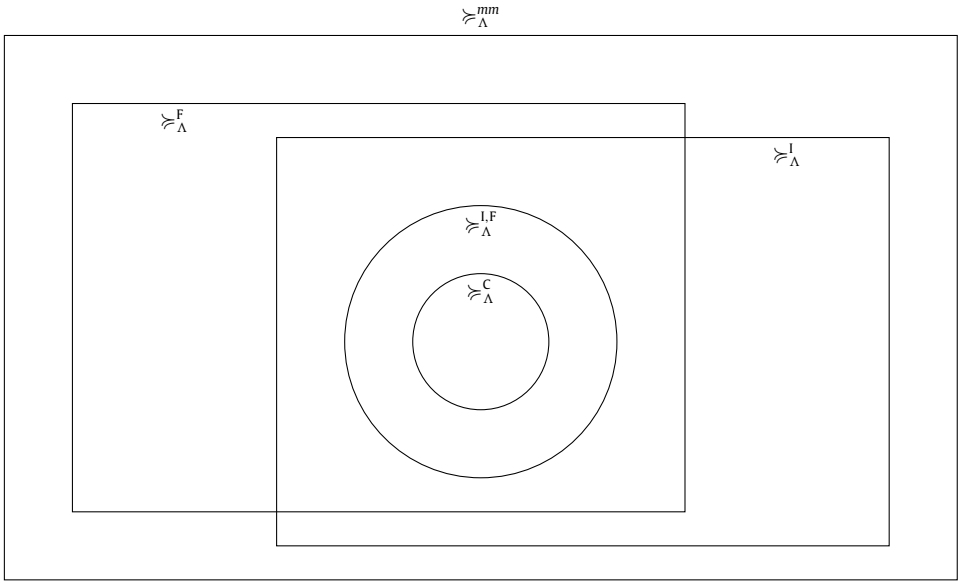


Fig. 4. The Venn diagram that depicts relationships between the preference relations defined in this paper.

7.2. Summary of computational characterisations

For finite consistent set of preference inputs $\Lambda \subseteq \mathbb{R}^n$ and arbitrary $\alpha, \beta \in \mathbb{R}^n$, we would like to be able to determine which of the following hold: $\alpha \succcurlyeq^{C}_{\Lambda} \beta$, $\alpha \succcurlyeq^{I}_{\Lambda} \beta$, $\alpha \succcurlyeq^{F}_{\Lambda} \beta$ and $\alpha \succcurlyeq^{I,F}_{\Lambda} \beta$. As usual, we label Λ as $\{\lambda_i : i \in I\}$. We use the results of previous sections to express, in terms of constraints, the condition that α does not dominate β, with respect to each of the four relations.

$\succcurlyeq^{C}_{\Lambda}$: $\alpha \not\succcurlyeq^{C}_{\Lambda} \beta$ if and only if, by Proposition 1(ii), there exists $u \in \Lambda^{\geq}$ such that $u \cdot \beta > u \cdot \alpha$. This holds if and only if there exists $u \in \mathbb{R}^n$, such that
- $u \cdot (\beta - \alpha) > 0$ and
- $\forall i \in I$, $u \cdot \lambda_i \geq 1$.

$\succcurlyeq^{I}_{\Lambda}$: $\alpha \not\succcurlyeq^{I}_{\Lambda} \beta$ if and only if there exists $u \in \mathrm{SI}(\Lambda)$ such that $u \cdot \beta > u \cdot \alpha$. Recall that, by Proposition 9, this holds if and only if there exists $u \in \mathbb{R}^n$, and non-negative reals r_i for each $i \in I$, such that
- $u \cdot (\beta - \alpha) > 0$;
- $\forall i \in I$, $u \cdot \lambda_i \geq 1$; and
- $u = \sum_{i \in I} r_i \lambda_i$.

Note that if $\mathbf{t}$ was not restricted to $(0, 1]^m$ in the definition of $\mathrm{SI}(\Lambda)$, then the second constraint (i.e., $u \cdot \lambda_i \geq 1$) would be replaced by $u \cdot \lambda_i > 0$ which is computationally more expensive due to the strict inequality. However, as we proved in Proposition 3, the result for both cases is the same.

$\succcurlyeq^{F}_{\Lambda}$: $\alpha \not\succcurlyeq^{F}_{\Lambda} \beta$ if and only if there exists $u \in \mathrm{SF}(\Lambda)$ such that $u \cdot \beta > u \cdot \alpha$. As we saw in Proposition 24, this holds if and only if there exists $u \in \mathbb{R}^n$ and $\mu \in \mathbb{R}^n$, and non-negative reals r_i for each $i \in I$, such that
- $u \cdot (\beta - \alpha) > 0$;
- $\forall i \in I$, $u \cdot \lambda_i \geq 1$;
- $\forall i \in I$, $[u \cdot \lambda_i = 1$ or $r_i = 0]$;
- $\mu = \sum_{i \in I} r_i \lambda_i$;
- $\sum_{i \in I} (r_i \neq 0) \leq n + 1$; and
- $\forall j \in \{1, \ldots, n\}$, $u(j) = 0 \iff \mu(j) = 0$, and $u(j) > 0 \iff \mu(j) > 0$.

In CPLEX, a disjunctive constraint such as $[w \cdot \lambda_i = 1$ or $r_i = 0]$ can be expressed as $(w \cdot \lambda_i == 1) + (r_i == 0) \geq 1$ (each logical proposition is treated as an integer; 0 for false and 1 for true).

$\succcurlyeq^{I,F}_{\Lambda}$: $\alpha \not\succcurlyeq^{I,F}_{\Lambda} \beta$ if and only if there exists $u \in \mathrm{SIF}(\Lambda)$ such that $u \cdot \beta > u \cdot \alpha$. Recall from Proposition 26 that this holds if and only if there exists $u \in \mathbb{R}^n$ and $\mu \in \mathbb{R}^n$, and non-negative reals r_i for each $i \in I$, such that
- $u \cdot (\beta - \alpha) > 0$;
- $\forall i \in I$, $u \cdot \lambda_i \geq 1$;
- $\mu = \sum_{i \in I} r_i \lambda_i$; and
- $\forall j \in \{1, \ldots, n\}$, $u(j) = 0 \iff \mu(j) = 0$, and $u(j) > 0 \iff \mu(j) > 0$.

8. Optimality operators

In many decision-making situations, there is no clear ordering on decisions (alternatives). There can often be a set of different scenarios with a different ordering on alternatives in each scenario. For example, for different scalings of preference inputs, we may have different orderings over a set of alternatives. In such a setup there are a number of natural ways of defining the set of optimal solutions (best alternatives or top recommended solutions).

We consider here two kinds of *optimality operators* in the sense of [52]; namely the set of *undominated* solutions, which is a natural generalisation of the Pareto-optimal set; and the set of *possibly optimal* solutions. The set of possibly optimal alternatives has been considered in a number of different situations, including for voting rules [53], for soft constraint optimisation [54], and for multi-objective optimisation [55,52].

Let $\succcurlyeq_S$ be any of the relations $\succcurlyeq^{\mathrm{C}}_{\Lambda}, \succcurlyeq^{\mathrm{I}}_{\Lambda}, \succcurlyeq^{\mathrm{F}}_{\Lambda}$ and $\succcurlyeq^{\mathrm{I,F}}_{\Lambda}$, where S is the corresponding set of scenarios for each relation (see Section 7.1), which are respectively $\mathrm{SC}(\Lambda)$ $(= \Lambda^{\geq})$, $\mathrm{SI}(\Lambda)$, $\mathrm{SF}(\Lambda)$ and $\mathrm{SIF}(\Lambda)$. We have then $\alpha \succcurlyeq_S \beta$ if and only if, for all $u \in S$, $u \cdot \alpha \geq u \cdot \beta$. We define $\succ_S$ to be the strict part of $\succcurlyeq_S$, so that $\alpha \succ_S \beta$ if and only if $\alpha \succcurlyeq_S \beta$ and $\beta \not\succcurlyeq_S \alpha$.

An alternative α is defined to be an element of $\mathbb{R}^n$. For a given finite set of alternatives A, the two optimality operators are defined as follows:

***UND*$_S$(*A*)** $(= UND_{\succ_S}(A))$ is the set of undominated elements with respect to relation $\succ_S$, i.e., $\alpha \in UND_S(A)$ if and only if there is no $\beta \in A$ such that $\beta \succ_S \alpha$.

***PO*$_S$(*A*)** is the set of elements that are optimal in some scenario. Thus, $\alpha \in PO_S(A)$ if and only if there exists $u \in S$ such that for all $\beta \in A$, $\alpha \cdot u \geq \beta \cdot u$. Elements of $PO_S(A)$ are said to be *possibly optimal* (in A, given S).

Algorithm 2 Finding Undominated Elements ($UND_S(A)$) Incrementally.

```
1: function INCREMENTAL-UNDOMINATED(A)
2:     Ω = {}                                   ▷ This set contains the undominated elements found so far.
3:     for all α ∈ A do
4:         ********** Stage one **********
5:         for all ω ∈ Ω do
6:             if ω ≽_S α & α ⋡_S ω then        ▷ α is dominated by ω?
7:                 go to 21                     ▷ Proceed to the next α.
8:             end if
9:         end for
10:        ************************************************************
11:        Reaching this point means α is not dominated by any element of Ω.
12:        Now, we will eliminate any ω in Ω that is dominated by α.
13:        ********** Stage two **********
14:        for all ω ∈ Ω do
15:            if α ≽_S ω & ω ⋡_S α then         ▷ α dominates ω?
16:                Ω = Ω − {ω}
17:            end if
18:        end for
19:        ************************************************************
20:        Ω = Ω + {α}
21:    end for
22:    return Ω
23: end function
```

From the definition of $UND_S(A)$ we have $\alpha \in UND_S(A)$ if and only if there is no $\beta \in A$ such that $\beta \succcurlyeq_S \alpha$ and $\alpha \not\succcurlyeq_S \beta$. Thus, in order to compute $UND_S(A)$ we can make use of the computation methods proposed in Section 7 for computing $\succcurlyeq_S$.

In contrast, $PO_S(A)$ cannot be computed just from $\succcurlyeq_S$, because $PO_S(A)$ is not a function of $\succcurlyeq_S$ but rather a function of S (i.e., SF, SI etc.). It follows immediately from the definition that for any finite set of alternatives A, $PO_S(A) = \bigcup_{u \in S} PO_{\{u\}}(A)$, and $PO_{\{u\}}(A)$ is the set of elements α of A maximising $u \cdot \alpha$. This in turn implies that for any subsets S and S' of $\mathbb{R}^n$, $PO_{S \cup S'}(A) = PO_S(A) \cup PO_{S'}(A)$, and if $S \subseteq S'$ then $PO_S(A) \subseteq PO_{S'}(A)$. Proposition 28 then leads easily to the following result, showing nestings of the different forms of Possibly Optimal set.

Proposition 29. *For any finite non-empty set $A \subseteq \mathbb{R}^n$ of alternatives, and any finite consistent set of preference inputs $\Lambda \subseteq \mathbb{R}^n$, $PO_{\mathrm{SF}(\Lambda)}(A) \cap PO_{\mathrm{SI}(\Lambda)}(A)$ is non-empty, and $PO_{\mathrm{SF}(\Lambda)}(A) \cup PO_{\mathrm{SI}(\Lambda)}(A) = PO_{\mathrm{SI}(\Lambda) \cup \mathrm{SF}(\Lambda)}(A) \subseteq PO_{\mathrm{SIF}(\Lambda)}(A) \subseteq PO_{\mathrm{SC}(\Lambda)}(A)$.*

For each of the sets S of scenarios $\mathrm{SC}(\Lambda)$ $\mathrm{SI}(\Lambda)$, $\mathrm{SF}(\Lambda)$ and $\mathrm{SIF}(\Lambda)$, we have a characterisation of the condition $[u \in S]$ in terms of constraints, see Proposition 1(ii), Corollary 8, Theorem 23 and Theorem 25, respectively. (Each corresponds to the set of constraints for $\alpha \not\succcurlyeq_S \beta$ shown for the associated relation in Section 7.2, omitting the first constraint $u \cdot (\beta - \alpha) > 0$.) We then define $\mathcal{C}_S(A, \alpha)$ to be this set of constraints plus the constraints: for all $\beta \in A$, $\alpha \cdot u \geq \beta \cdot u$. Hence, u is a solution of $\mathcal{C}_S(A, \alpha)$ if and only if $u \in S$ and for all $\beta \in A$, $\alpha \cdot u \geq \beta \cdot u$. Therefore $\alpha \in PO_S(A)$ if and only if $\mathcal{C}_S(A, \alpha)$ has a solution.

For example, $\mathcal{C}_{\mathrm{SI}(\Lambda)}(A,\alpha)$ is the following set of constraints (compare this with $\succcurlyeq^{\mathrm{I}}_{\Lambda}$ in Section 7):

- $\forall \beta \in A,\ u \cdot (\alpha - \beta) \geq 0$;
- $\forall i \in I,\ u \cdot \lambda_i \geq 1$; and
- $u = \sum_{i \in I} r_i \lambda_i$.

Typically, but not always (as we found in our experiments), $PO_S(A)$ is a smaller set than $UND_S(A)$ (since possibly optimal alternatives are very often, but not always, undominated).

Propositions 2 and 4 in [52] imply that computation of $UND_S(A)$ and $PO_S(A)$ can be done with a very simple incremental algorithm. We adapt this incremental approach and exploit it for each of the four sets of scenarios.

Algorithm 2 shows how $UND_S(A)$ can be found incrementally. It corresponds with a natural way of computing Pareto optimal solutions. The algorithm consists of two stages for each $\alpha \in A$. In the first stage, we examine if α is undominated among the undominated elements Ω found so far. We proceed to the next stage if α is undominated and remove those elements of Ω that are dominated by α (so they are no longer undominated). The correctness of Algorithm 2 is formally stated in Proposition 30.

Proposition 30. *For finite consistent set of preference inputs $\Lambda \subseteq \mathbb{R}^n$, given any subset S of $\Lambda^{\geq}$, and any finite set A ($\subseteq \mathbb{R}^n$) of alternatives, Algorithm 2 returns $UND_S(A)$.*

Algorithm 3 Finding Possibly Optimal Elements ($PO_S(A)$) Incrementally.

```
1: function INCREMENTAL-PO(A)
2:     Ψ = {}
3:     for all α ∈ A do
4:         u = F_S(Ψ↓, α)                          ▷ Solving for α against POs that are found so far.
5:         if u is not NULL then                   ▷ α is a PO element within Ψ↓?
6:             Ψ = Ψ + {(α, u)}                    ▷ Keep also the solution u for later use.
7:             REFINE-PREVIOUS-POS(Ψ, α)
8:         end if
9:     end for
10:    return Ψ↓
11: end function
12: ************************************************************************************
13: The following function eliminates ψ ∈ Ψ↓ which are no longer PO due to arrival of α.
14: ************************************************************************************
15: function REFINE-PREVIOUS-POS(Ψ, α)
16:    for all (ψ, v) ∈ Ψ do
17:        if ψ · v < α · v then                   ▷ Is α better than ψ in the scenario v in which ψ was optimal?
18:            Ψ = Ψ − {(ψ, v)}
19:            u = F_S(Ψ↓, ψ)                      ▷ Check if there is another solution which makes ψ optimal.
20:            if u is not NULL then
21:                Ψ = Ψ + {(ψ, u)}                ▷ Add ψ again but with the new solution u (rather than v).
22:            end if
23:        end if
24:    end for
25: end function
```

The set of possibly optimal elements $PO_S(A)$ is built up in an incremental way in Algorithm 3. In this algorithm, $\mathcal{F}_S(A,\alpha)$ is a function such that it returns the solution of $\mathcal{C}_S(A,\alpha)$ if a solution is found, and *NULL* otherwise. Here, Ψ is a set of pairs where the first component of a pair is the potentially possibly optimal element, and the second one is the scenario in which the first component has been found to be optimal. Regarding this notation, $\Psi^{\downarrow}$ is the set of first components in Ψ; i.e., $\Psi^{\downarrow} = \{\psi : (\psi, u) \in \Psi\}$. In Line 6, once it is found out that α is a possibly optimal element within $\Psi^{\downarrow}$, it is included in Ψ along with its associated solution (scenario). Then, in the function *Refine-Previous-POs*, we remove any $(\psi, v) \in \Psi$ which is not possibly optimal anymore because of adding α. In Line 18, the existing possibly optimal element ψ is removed from Ψ because it is not as good as the incoming possibly optimal element α in its own associated scenario v. However, it does not mean that ψ cannot be possibly optimal; there might be another scenario u in which ψ is better than all elements of Ψ including α. If it is the case, we include ψ again in Ψ but with this new scenario u instead of v. Proposition 31 formally states the correctness of Algorithm 3.

Proposition 31. *For finite consistent set of preference inputs $\Lambda \subseteq \mathbb{R}^n$, given any subset S of $\Lambda^{\geq}$, and any finite set A ($\subseteq \mathbb{R}^n$) of alternatives, Algorithm 3 returns $PO_S(A)$.*

Table 2
The results relate to determining decisive pairs, among 1000 pairs of test vectors with respect to preference relations $\succcurlyeq^{F}_{\Lambda}$, $\succcurlyeq^{I}_{\Lambda}$, the intersection of $\succcurlyeq^{I}_{\Lambda}$ and $\succcurlyeq^{F}_{\Lambda}$ ($\succcurlyeq^{I\wedge F}_{\Lambda}$), $\succcurlyeq^{I,F}_{\Lambda}$, and $\succcurlyeq^{C}_{\Lambda}$. The bold numbers indicate that it is not always the case that $\succcurlyeq^{F}_{\Lambda}$ is a weaker relation than $\succcurlyeq^{I}_{\Lambda}$. The last row includes the mean of the values of each column, rounded to the nearest integer.

	m	Decisive Pairs (%)					Time (ms)			
		$\succcurlyeq^{F}_{\Lambda}$	$\succcurlyeq^{I}_{\Lambda}$	$\succcurlyeq^{I\wedge F}_{\Lambda}$	$\succcurlyeq^{I,F}_{\Lambda}$	$\succcurlyeq^{C}_{\Lambda}$	$\succcurlyeq^{F}_{\Lambda}$	$\succcurlyeq^{I}_{\Lambda}$	$\succcurlyeq^{I,F}_{\Lambda}$	$\succcurlyeq^{C}_{\Lambda}$
Ridesharing Database										
1.	24	21	16	9	3	1	517	36	55	18
2.	29	92	31	31	26	0.3	2434	23	40	16
3.	31	**23**	**28**	13	1	0.1	800	25	38	13
4.	36	81	35	35	31	23	4768	24	43	14
5.	38	36	19	17	5	2	2799	24	47	17
6.	41	61	12	12	12	12	5123	23	45	20
7.	53	40	20	19	19	19	1134	24	41	20
8.	55	97	26	26	24	8	1833	26	45	19
9.	62	48	24	24	11	1	4983	27	50	14
10.	94	64	35	35	5	2	5084	27	54	23
11.	127	62	24	24	24	13	6439	28	57	21
12.	129	80	36	36	19	1	2928	30	49	25
13.	134	69	28	28	28	16	7374	30	48	19
Mean	66	59	26	24	16	8	3555	27	48	19
Car Preference Database										
1.	26	65	32	31	22	10	2731	28	64	27
2.	30	42	36	28	21	12	1962	26	94	23
3.	30	36	19	17	11	7	4700	23	56	25
4.	35	56	33	30	22	9	6612	24	149	22
5.	35	65	18	17	11	5	5850	25	77	22
6.	35	**49**	**61**	41	34	20	377	26	72	31
7.	40	53	36	33	24	13	1411	56	173	78
8.	40	68	46	46	34	15	2879	26	64	25
9.	40	42	39	28	21	14	1150	26	78	27
10.	41	51	35	29	24	12	1317	28	97	23
Mean	35	53	35	30	22	12	2899	29	93	30

9. Experimental testing

In this section we experimentally test the methods and algorithms developed in earlier sections; the results show the feasibility of the methods, and illustrate relative computational efficiency, as well as the differences between the various relations and optimality classes. It is shown that our preference relations do not necessarily lead to a large number of solutions for the decision maker to consider.

The experiments make use of two databases, namely *Ridesharing Database* and *Car Preference Database*. The ridesharing database is a subset of a year's worth of real ridesharing records, provided by a commercial ridesharing system *Carma* (see http://gocarma.com/). Each ridesharing alternative has 7 features, representing different aspects of a possible choice of match for a given user. More information about the data can be found in [45].

The second database is the result of a survey expressing the preferences of different users over specific cars [56]. For each car 7 features are considered (e.g., engine size).

We base our experiments on 13 benchmarks derived from the ridesharing database and 10 benchmarks derived from the car preference database. Each benchmark corresponds to the inferred preferences of a different user. The preference of alternative a_i (i.e., a ridesharing alternative or a car) over b_i leads to $a_i - b_i (= \lambda_i)$ being included in Λ.

A pre-processing phase deletes some elements of Λ, in order to make it consistent (i.e., $\Lambda^{\geq} \neq \emptyset$). In order to do that, we adopt the first and the second approaches discussed in Section 6 respectively for the first and the second database. To conduct the experiments, CPLEX 12.6.3 is used as the solver on a computer facilitated by an Intel Xeon E312xx 2.20 GHz processor and 8 GB RAM memory.

9.1. Decisive pairs

Here, we would like to examine how decisive each relation is, i.e., which relation is weaker and by how much. We randomly generate 1000 pairs (α, β), based on a uniform distribution for each feature. A pair (α, β) is called *decisive* for a preference relation if one of them can (strictly) dominate the other one; for example, the pair (α, β) is decisive for $\succcurlyeq^{I}_{\Lambda}$ if and only if $\alpha \succ^{I}_{\Lambda} \beta$ or $\beta \succ^{I}_{\Lambda} \alpha$. This is if and only if either ($\alpha \succcurlyeq^{I}_{\Lambda} \beta$ and $\beta \not\succcurlyeq^{I}_{\Lambda} \alpha$) or ($\beta \succcurlyeq^{I}_{\Lambda} \alpha$ and $\alpha \not\succcurlyeq^{I}_{\Lambda} \beta$). We also consider the relation $\succcurlyeq^{I\wedge F}_{\Lambda}$ which is the intersection of $\succcurlyeq^{I}_{\Lambda}$ and $\succcurlyeq^{F}_{\Lambda}$ (see Section 7.1; note that this relation differs from the relation $\succcurlyeq^{I,F}_{\Lambda}$).

Table 3
A comparison, between the number of possibly optimal elements and the number of undominated elements among 100 alternatives with regard to preference relations $\succcurlyeq^{C}_{\Lambda}$, $\succcurlyeq^{I,F}_{\Lambda}$, $\succcurlyeq^{I}_{\Lambda}$, and $\succcurlyeq^{F}_{\Lambda}$. The I ∩ F column relates to the intersection of the I and F columns. The bold numbers illustrate that the F and I ∩ F sets are not always identical (so that the F optimality set is not always a subset of the I optimality set), and the encircled numbers relate to the cases when $|PO_S(A)| > |UND_S(A)|$. The last row includes the mean of values of each column, rounded to the nearest integer.

	$\lvert PO_S(A)\rvert$					$\lvert UND_S(A)\rvert$				
	C	I, F	I	F	I ∩ F	C	I, F	I	F	I ∩ F
Ridesharing Database										
1.	38	26	20	**6**	**4**	72	55	33	**16**	**13**
2.	45	13	12	2	2	86	20	15	3	3
3.	64	37	21	**6**	**5**	97	74	30	**19**	**18**
4.	7	7	7	3	3	7	7	7	4	4
5.	33	32	21	**13**	**12**	63	54	38	17	17
6.	14	14	14	5	5	18	18	18	5	5
7.	10	10	10	6	6	18	18	17	7	7
8.	18	9	9	1	1	25	12	12	1	1
9.	34	17	13	6	6	78	19	15	8	8
10.	22	15	8	2	2	50	38	13	2	2
11.	20	14	14	2	2	27	19	19	3	3
12.	41	12	9	2	2	79	24	15	2	2
13.	16	12	12	4	4	29	16	16	6	6
Mean	28	17	13	4	4	50	29	19	7	7
Car Preference Database										
1.	14	14	11	**6**	**5**	18	18	12	6	6
2.	14	10	3	**5**	**2**	15	11	4	**5**	**3**
3.	34	30	16	**7**	**3**	36	31	26	12	12
4.	17	11	8	2	2	20	13	11	3	3
5.	29	22	21	(11)	(11)	34	26	21	(9)	(7)
6.	8	5	3	2	2	8	5	3	2	2
7.	12	6	5	3	3	12	6	5	3	3
8.	14	11	5	**4**	**2**	15	13	8	**4**	**2**
9.	14	11	7	(5)	(4)	14	11	7	(4)	(3)
10.	15	10	9	4	4	16	10	9	4	4
Mean	17	13	9	5	4	19	14	11	5	5

To determine whether a pair is decisive we need to run the solver, based on the proposed computation methods in Section 7, twice; once for testing if $\alpha \succcurlyeq_S \beta$ and a second time for $\beta \succcurlyeq_S \alpha$.

Table 2 shows the percentage of decisive pairs for $\succcurlyeq^{F}_{\Lambda}$, $\succcurlyeq^{I}_{\Lambda}$, $\succcurlyeq^{I\wedge F}_{\Lambda}$, $\succcurlyeq^{I,F}_{\Lambda}$ and $\succcurlyeq^{C}_{\Lambda}$, as well as the running time per pair. The results illustrate some of the relationships expressed in Proposition 28: $\succcurlyeq^{C}_{\Lambda} \subseteq \succcurlyeq^{I,F}_{\Lambda} \subseteq \succcurlyeq^{I\wedge F}_{\Lambda}$ (which equals $\succcurlyeq^{I}_{\Lambda} \cap \succcurlyeq^{F}_{\Lambda}$). They also demonstrate that the subset relations can easily be strict, with $\succcurlyeq^{I,F}_{\Lambda}$ not being the same as either the relation $\succcurlyeq^{I\wedge F}_{\Lambda}$ or the consistency-based relation $\succcurlyeq^{C}_{\Lambda}$. Typically, relation $\succcurlyeq^{F}_{\Lambda}$, relating to rescaling of the features, is much the strongest relation, i.e., most decisive, followed by $\succcurlyeq^{I}_{\Lambda}$, which is only slightly more decisive than $\succcurlyeq^{I\wedge F}_{\Lambda}$, which is a good deal stronger than $\succcurlyeq^{I,F}_{\Lambda}$, with the consistency-based relation $\succcurlyeq^{C}_{\Lambda}$ being much the weakest (least decisive).

The fact that the relation $\succcurlyeq^{I}_{\Lambda}$, based on preference inputs rescaling, is only slightly more decisive than $\succcurlyeq^{I\wedge F}_{\Lambda}$ suggests that $\succcurlyeq^{I}_{\Lambda}$ can be close (in some sense) to being a sub-relation of $\succcurlyeq^{F}_{\Lambda}$, since if $\succcurlyeq^{I}_{\Lambda} \subseteq \succcurlyeq^{F}_{\Lambda}$ then $\succcurlyeq^{I\wedge F}_{\Lambda} = \succcurlyeq^{I}_{\Lambda}$. However, in four of the thirteen ridesharing benchmarks, and in nine of the ten Car Preference benchmarks (see Table 2), the number of decisive pairs for $\succcurlyeq^{I}_{\Lambda}$ is not equal to the number for $\succcurlyeq^{I\wedge F}_{\Lambda}$. This implies that in these particular benchmarks, we have $\succcurlyeq^{F}_{\Lambda} \not\supseteq \succcurlyeq^{I}_{\Lambda}$ (and hence $SF(\Lambda) \not\subseteq SI(\Lambda)$). There are even two of the benchmarks (see the figures in bold) in which $\succcurlyeq^{F}_{\Lambda}$ is less decisive than $\succcurlyeq^{I}_{\Lambda}$.

In terms of running time, $\succcurlyeq^{I}_{\Lambda}$ is around 130 and 100 times faster than $\succcurlyeq^{F}_{\Lambda}$ on average for the ridesharing database and the car preference database, respectively. The computations for $\succcurlyeq^{C}_{\Lambda}$ and $\succcurlyeq^{I,F}_{\Lambda}$ are of the same order of magnitude as for $\succcurlyeq^{I}_{\Lambda}$, with the former being somewhat faster for the ridesharing database, and those for $\succcurlyeq^{I,F}_{\Lambda}$ being somewhat slower. It is interesting that the non-linear constraint for $\succcurlyeq^{I,F}_{\Lambda}$ (see Section 7.2) makes much less of a difference for computation time than the non-linear constraints for computing $\succcurlyeq^{F}_{\Lambda}$. The computation times do not appear to depend strongly on the number m of preference inputs, with the partial exception of the $\succcurlyeq^{F}_{\Lambda}$ relation.

Table 4
A comparison, between the running time for finding possibly optimal elements and undominated elements among 100 alternatives with regard to preference relations $\succcurlyeq_{\Lambda}^{C}$, $\succcurlyeq_{\Lambda}^{I,F}$, $\succcurlyeq_{\Lambda}^{I}$ and $\succcurlyeq_{\Lambda}^{F}$. The last row includes the mean of values of each column, rounded to the nearest integer.

	$PO_S(A)$ Time (s)				$UND_S(A)$ Time (s)			
	C	I, F	I	F	C	I, F	I	F
Ridesharing Database								
1.	31	53	18	66	215	187	97	128
2.	41	39	22	505	152	46	24	516
3.	37	103	17	234	176	241	43	570
4.	7	11	13	29	9	18	10	870
5.	13	29	22	572	124	166	68	1710
6.	17	22	21	277	32	53	29	1723
7.	11	14	17	86	24	32	21	259
8.	13	16	8	334	42	20	12	243
9.	26	34	9	405	162	42	23	1729
10.	27	31	14	654	151	136	25	1147
11.	15	27	13	412	51	48	33	1835
12.	41	23	14	330	272	46	22	558
13.	27	24	19	539	68	46	29	2087
Mean	24	33	16	342	114	83	34	1029
Car Preference Database								
1.	14	62	21	226	37	88	19	471
2.	10	47	8	185	25	38	11	870
3.	18	33	13	407	51	91	53	2344
4.	10	51	15	338	19	42	13	2126
5.	22	35	17	603	43	156	33	1535
6.	11	30	13	40	35	44	53	367
7.	12	48	20	98	63	63	55	275
8.	10	52	10	294	20	22	8	415
9.	10	24	13	55	22	50	6	72
10.	7	34	10	149	26	41	9	374
Mean	12	42	14	240	34	64	26	885

9.2. Optimal elements

The next phase of experiments is devoted to finding optimal solutions with respect to the two kinds of optimality operator discussed in Section 8. To do so, a set of 100 alternatives (i.e., the set A) is randomly generated, based on a uniform distribution for each feature. Then, for each relation, the number of possibly optimal and undominated elements in A is counted; see Table 3. The numbers in the I ∩ F columns relate to the intersection of the I and F optimality sets; for example, the left-hand I ∩ F column gives the cardinalities of the sets $PO_{\mathrm{SI}(\Lambda)} \cap PO_{\mathrm{SF}(\Lambda)}$.

The results in Table 3 illustrate the following connections, stated in Proposition 29, between the different Possibly Optimal sets: $PO_{\mathrm{SF}(\Lambda)}(A) \cup PO_{\mathrm{SI}(\Lambda)}(A) = PO_{\mathrm{SI}(\Lambda)\cup\mathrm{SF}(\Lambda)}(A) \subseteq PO_{\mathrm{SIF}(\Lambda)}(A) \subseteq PO_{\mathrm{SC}(\Lambda)}(A)$. Although it is not given explicitly in Table 3, we can compute $|PO_{\mathrm{SI}(\Lambda)\cup\mathrm{SF}(\Lambda)}(A)|$ using $|PO_{\mathrm{SI}(\Lambda)\cup\mathrm{SF}(\Lambda)}(A)| = |PO_{\mathrm{SF}(\Lambda)}(A) \cup PO_{\mathrm{SI}(\Lambda)}(A)| = |PO_{\mathrm{SF}(\Lambda)}(A)| + |PO_{\mathrm{SI}(\Lambda)}(A)| - |PO_{\mathrm{SF}(\Lambda)}(A) \cap PO_{\mathrm{SI}(\Lambda)}(A)|$. This implies that in nine of the Ridesharing benchmarks and in all of the Car Preference benchmarks, $PO_{\mathrm{SIF}(\Lambda)}(A)$ is a strict superset of $PO_{\mathrm{SI}(\Lambda)\cup\mathrm{SF}(\Lambda)}(A)$.

The results also show that others of the subset relations can easily be strict. For example, with the first Ridesharing benchmark, $PO_{\mathrm{SF}(\Lambda)}(A) \cap PO_{\mathrm{SI}(\Lambda)}(A) \subsetneq PO_{\mathrm{SF}(\Lambda)}(A) \subsetneq PO_{\mathrm{SI}(\Lambda)\cup\mathrm{SF}(\Lambda)}(A) \subsetneq PO_{\mathrm{SIF}(\Lambda)}(A) \subsetneq PO_{\mathrm{SC}(\Lambda)}(A)$, and similarly, with $PO_{\mathrm{SI}(\Lambda)}(A)$ replacing $PO_{\mathrm{SF}(\Lambda)}(A)$.

In most of the benchmarks the figure in the I ∩ F column for the $PO_S(A)$ case is equal to the corresponding value in the F column, which implies that $PO_{\mathrm{SF}(\Lambda)}(A)$ is then a subset of $PO_{\mathrm{SI}(\Lambda)}(A)$, and similarly, for the $UND_S(A)$ results. However, the bold numbers show that the F and I ∩ F columns are not identical, and thus illustrate that e.g., $PO_{\mathrm{SF}(\Lambda)}(A)$ is not necessarily a subset of $PO_{\mathrm{SI}(\Lambda)}(A)$.

One can sometimes obtain a still smaller set than that related to SF(Λ) by taking the intersection of the optimality sets for SI(Λ) and SF(Λ). For the Possibly Optimal case, this set $PO_{\mathrm{SI}(\Lambda)} \cap PO_{\mathrm{SF}(\Lambda)}$ is guaranteed to be non-empty, by Proposition 29 (because it contains the non-empty set $PO_{\{\omega^{\star}_{\Lambda}\}}(A)$).

In the ridesharing database, it can be seen that for the most conservative relation, $\succcurlyeq_{\Lambda}^{C}$, the optimality operators return a substantial proportion of alternatives as optimal solutions (roughly half for $UND_S(A)$).

The results for SIF(Λ) (invariant to preference inputs and features rescaling), the most robust of the three rescaling approaches, lead to only slightly more optimal solutions than for SI(Λ). Also, for the ridesharing benchmarks, the $PO_S(A)$ sets tend to be substantially smaller than the corresponding $UND_S(A)$ sets. However, the number of undominated elements for the car preference database is fairly similar to the number of possibly optimal elements, and we sometimes even have $|PO_S(A)| > |UND_S(A)|$ (see the encircled numbers).

Table 4 shows the time for finding possibly optimal and undominated solutions, where the former is faster than the latter by a factor ranging from 1.5 to 4.8 on average; this is partly because of $|PO_S(A)|$ being usually smaller than $|UND_S(A)|$ particularly for the ridesharing database. Because the computation of $\succcurlyeq^{\mathrm{F}}_{\Lambda}$ was very much slower than the other relations, the times in the F columns are still greatest, despite the number of optimal solutions being smaller. Overall, the computational cost of the relation $\succcurlyeq^{\mathrm{F}}_{\Lambda}$ may make it less useful, even though it is more decisive, and thus leads to smaller sets of optimal solutions. Instead one might, for instance, favour $PO_{\mathrm{SI}(\Lambda)}$, $PO_{\mathrm{SIF}(\Lambda)}$ and $UND_{\mathrm{SI}(\Lambda)}$ since they generate reasonably sized optimality sets much faster. Recall that each one of the returned solutions in a Possibly Optimal set is an optimal solution to a rescaled version of the original problem; it thus seems natural for it to be available for consideration by the decision maker.

10. Summary and discussion

The maximum margin method for preference learning learns a utility function from a set of input preferences, in order to predict further preferences. However, in many situations, it can be argued that the scaling of preference inputs should not affect the induced preference relation. We have defined a relation $\succcurlyeq^{\mathrm{I}}_{\Lambda}$ that is a more robust version of the maximum margin preference inference $\succcurlyeq^{mm}_{\Lambda}$, and which is invariant to the scaling of preference inputs. It is also reasonable to consider invariance to the way that features are scaled because, in maximum margin inference, features should be scaled before applying the method; this is due to the fact that the objective function in maximum margin method is sensitive to the scale of feature domains. Thus, we have also defined the $\succcurlyeq^{\mathrm{F}}_{\Lambda}$ relation which is invariant to the scaling of features. With these two types of rescaling being complementary, it is also natural to consider both types simultaneously, leading to a further preference relation $\succcurlyeq^{\mathrm{I,F}}_{\Lambda}$. We derived characterisations for the relations $\succcurlyeq^{\mathrm{I}}_{\Lambda}$, $\succcurlyeq^{\mathrm{F}}_{\Lambda}$ and $\succcurlyeq^{\mathrm{I,F}}_{\Lambda}$, which lead to computational procedures. We also characterised the situation when the maximum margin relation is insensitive to the scaling of features, i.e., $\succcurlyeq^{\mathrm{F}}_{\Lambda}$ equals $\succcurlyeq^{mm}_{\Lambda}$. We then discussed three basic approaches to restore consistency of input data. Two optimality operators—$UND_S(A)$ and $PO_S(A)$—have been considered to define how a set of optimal solutions can be extracted from the available alternatives. We proposed two algorithms in order to compute $UND_S(A)$ and $PO_S(A)$ in an incremental manner. Our experiments, which used 23 benchmarks derived from two sets of real preference data, compared the different relations in terms of decisiveness and the set of optimal solutions regarding $UND_S(A)$ and $PO_S(A)$, and showed that the computational methods are practically feasible for a moderate number of instances/features. The relation associated with only scaling the features was the most decisive but by far the slowest for computing the associated optimality classes. Overall, one might consider $\succcurlyeq^{\mathrm{I}}_{\Lambda}$ as a relation that keeps quite a good balance between decisiveness and computation time.

In the future, it would be interesting to explore extensions of our approaches including (i) integration of the approach with a conversational recommender system, and with a multi-criteria decision-making system; (ii) developing computational methods for certain kinds of kernel; (iii) considering soft margin optimisation, i.e., more sophisticated approaches for dealing with an inconsistent dataset; (iv) taking into account more general kinds of input preference statement; and (v) exploring connections with imprecise probability, based on linear constraints on probabilities.

CRediT authorship contribution statement

Mojtaba Montazery: Conceptualization, Formal analysis, Investigation, Methodology, Software, Visualization, Writing - original draft. **Nic Wilson:** Conceptualization, Formal analysis, Investigation, Methodology, Supervision, Writing - original draft, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under Grant Numbers SFI/12/RC/2289 and 12/RC/2289-P2 which are co-funded under the European Regional Development Fund. We're grateful to the reviewers for their comments, which helped improve the final version of the paper.

Appendix

The appendix includes all the proofs, of the results in the paper, that do not appear in the main body of the paper.

Results in Section 2

The proof of Proposition 1 uses Lemmas 32 and 33.

Lemma 32. *Consider any $\Lambda \subseteq \mathbb{R}^n$. If $\Lambda^>$ is non-empty then Λ^* is the topological closure of $\Lambda^>$.*

Proof. Let us write the topological closure operator as $Cl(\cdot)$, so that $Cl(S)$ is the topological closure of S, which equals S plus all the limit points of S. Basic properties of $Cl(\cdot)$ include: (a) $S \subseteq T$ implies $Cl(S) \subseteq Cl(T)$, and (b) $Cl(S) = S$ if S is a topologically closed set.

It is clear that $\Lambda^> \subseteq \Lambda^*$ which implies that $Cl(\Lambda^>) \subseteq Cl(\Lambda^*)$; also, $Cl(\Lambda^*) = \Lambda^*$ since Λ^* is a topologically closed set. We thus have $Cl(\Lambda^>) \subseteq \Lambda^*$.

Now, we will show that $\Lambda^* \subseteq Cl(\Lambda^>)$. To do so, we will prove that for any $x \in \Lambda^*$ there is a sequence of elements of $\Lambda^>$ that converges to x. Choose arbitrary $x \in \Lambda^*$ and $y \in \Lambda^>$. For each $\lambda \in \Lambda$, $x \cdot \lambda \geq 0$ and $y \cdot \lambda > 0$, and thus, for each $\delta \in (0,1)$, $(\delta x + (1-\delta)y) \cdot \lambda = \delta(x \cdot \lambda) + (1-\delta)(y \cdot \lambda) > 0$, and so, $\delta x + (1-\delta)y \in \Lambda^>$. As δ tends to 1, $\delta x + (1-\delta)y$ tends to x, showing that $x \in Cl(\Lambda^>)$, as required. □

The following lemma is a well-known result for convex cones. Consider any finite $\Lambda \subseteq \mathbb{R}^n$ and any $u \in \mathbb{R}^n$. Then, $\Lambda^* \subseteq \{u\}^*$ if and only if $u \in co(\Lambda)$.

Lemma 33. *Consider any finite $\Lambda \subseteq \mathbb{R}^n$ and any $u \in \mathbb{R}^n$. Then, $\Lambda^* \subseteq \{u\}^*$ if and only if $u \in co(\Lambda)$.*

To illustrate, $u = (3,3)$ is in $co(\Lambda)$ in Fig. 1, and $\{u\}^* = \{(x,y) : 3x + 3y \geq 0\}$ clearly contains $\Lambda^{\geq}$, the union of the shaded regions in Fig. 1(b).

Proof. Because $\Lambda^* = (co(\Lambda))^*$ we have that $\Lambda^* \subseteq (\{u\})^*$ if and only if $(co(\Lambda))^* \subseteq (co(\{u\}))^*$. Now, clearly, if $u \in co(\Lambda)$ then $(co(\Lambda))^* \subseteq (\{u\})^*$, and thus $\Lambda^* \subseteq (\{u\})^*$. To prove the converse, it is sufficient to show that $(co(\Lambda))^* \subseteq (co(\{u\}))^*$ implies $u \in co(\Lambda)$. Now, $(co(\Lambda))^* \subseteq (co(\{u\}))^*$ implies $(co(\Lambda))^{**} \supseteq (co(\{u\}))^{**}$. Convex cones $co(\Lambda)$ and $co(\{u\})$ are both topologically closed (because Λ is finite), so, by a fundamental result for convex cones $(co(\Lambda))^{**} = co(\Lambda)$ and $(co(\{u\}))^{**} = co(\{u\})$, and thus $co(\{u\}) \subseteq co(\Lambda)$, which implies that $u \in co(\Lambda)$. □

Proposition 1. *Consider any finite $\Lambda \subseteq \mathbb{R}^n$ that is consistent (i.e., $\Lambda^> \neq \emptyset$) and consider any $\gamma \in \mathbb{R}^n$. Then, the following conditions are equivalent. Thus, any of these are equivalent to $\gamma \succcurlyeq^C_\Lambda \vec{\mathbf{0}}$.*

(i) *for all $w \in \Lambda^>$, $w \cdot \gamma \geq 0$.*
(ii) *for all $w \in \Lambda^{\geq}$, $w \cdot \gamma \geq 0$.*
(iii) *$\gamma \in co(\Lambda)$.*

Proof. (i) ⇒ (ii): This follows immediately from $\Lambda^{\geq} \subset \Lambda^>$.

(i) ⇐ (ii): Suppose that for all $w \in \Lambda^{\geq}$, $w \cdot \gamma \geq 0$, and consider any $u \in \Lambda^>$. Let $a_u = \min_{\lambda \in \Lambda} u \cdot \lambda$ which is clearly greater than zero, and let $u' = \frac{u}{a_u}$. For any $\lambda \in \Lambda$, $u \cdot \lambda \geq a_u$ which implies that $u' \cdot \lambda \geq 1$, and thus, $u' \in \Lambda^{\geq}$. Because $u' \cdot \gamma \geq 0$, we have also, $u \cdot \gamma \geq 0$.

(i) ⇔ (iii): (i) means $\Lambda^> \subseteq \{\gamma\}^*$ which, because $\{\gamma\}^*$ is a closed set, holds if and only if $Cl(\Lambda^>) \subseteq \{\gamma\}^*$, i.e., $\Lambda^* \subseteq \{\gamma\}^*$, using Lemma 32. Lemma 33 implies that this is if and only if (iii) $\gamma \in co(\Lambda)$. □

The proof of Theorem 2 uses the following lemma.

Lemma 34. *For $w \in \Lambda^>$, define a_w to be $\min_{\lambda \in \Lambda} w \cdot \lambda$, (which is always strictly positive by definition of $\Lambda^>$), and define $\bar{w}$ to be $\frac{w}{a_w}$. Then, the following hold for any $w \in \Lambda^>$. (i) $\bar{w} \in \Lambda^{\geq}$; (ii) if $w \in \Lambda^{\geq}$ then $\|w\| > \|\bar{w}\|$ unless $w = \bar{w}$; (iii) for any real $r > 0$, $\mathrm{marg}_\Lambda(rw) = \mathrm{marg}_\Lambda(w)$; (iv) $\mathrm{marg}_\Lambda(w) = \frac{1}{\|\bar{w}\|}$.*

Proof. Assume $w \in \Lambda^>$. Then $a_{\bar{w}} = \min_{\lambda \in \Lambda} \frac{1}{a_w} w \cdot \lambda = \frac{a_w}{a_w} = 1$. Thus, $\bar{w} \in \Lambda^{\geq}$, showing (i). Also, $\frac{\|w\|}{\|\bar{w}\|} = a_w$, by definition of $\bar{w}$. If $w \in \Lambda^{\geq}$ then $a_w \geq 1$, so $\|w\| > \|\bar{w}\|$ unless $a_w = 1$, i.e., $w = \bar{w}$, proving (ii). The definitions immediately imply that $\mathrm{marg}_\Lambda(w) = \frac{a_w}{\|w\|}$. Since $a_{\bar{w}} = 1$, we have $\mathrm{marg}_\Lambda(\bar{w}) = \frac{1}{\|\bar{w}\|}$. The definition of marg_Λ implies that for any real $r > 0$, $\mathrm{marg}_\Lambda(rw) = \mathrm{marg}_\Lambda(w)$, showing (iii), so, in particular, $\mathrm{marg}_\Lambda(w) = \mathrm{marg}_\Lambda(\bar{w}) = \frac{1}{\|\bar{w}\|}$, which proves (iv). □

Theorem 2. *Let $\Lambda \subseteq \mathbb{R}^n$ be a finite consistent set of preference inputs, so that $\Lambda^>$ is non-empty. Then the following all hold.*

(i) *$\Lambda^{\geq}$ is non-empty;*
(ii) *there exists a unique element $\omega^\star_\Lambda$ in $\Lambda^{\geq}$ with minimum norm;*
(iii) *w maximises marg_Λ within $\Lambda^>$ if and only if w is a strictly positive scalar multiple of $\omega^\star_\Lambda$, i.e., there exists $r \in \mathbb{R}$ with $r > 0$ such that $w = r\omega^\star_\Lambda$.*

Proof. If $\Lambda^{>}$ is non-empty then, by Lemma 34(i), $\bar{w} \in \Lambda^{\geq}$ for any $w \in \Lambda^{>}$. Thus, $\Lambda^{\geq}$ is non-empty, showing (i). Regarding (ii), since $\Lambda^{\geq}$ is convex and topologically closed, there exists a unique element $\omega^{\star}_{\Lambda}$ in $\Lambda^{\geq}$ with minimum norm, by a standard result (for a proof, see e.g., Proposition 4 of [49]).

To prove (iii), consider any $w \in \Lambda^{>}$. As we just showed, $\bar{w} \in \Lambda^{\geq}$, so minimality of $\omega^{\star}_{\Lambda}$ implies that $\|\bar{w}\| \geq \|\omega^{\star}_{\Lambda}\|$ which equals $\|\overline{\omega^{\star}_{\Lambda}}\|$, using Lemma 34(ii). Lemma 34(iv) then implies that $\text{marg}_{\Lambda}(w) \leq \text{marg}_{\Lambda}(\omega^{\star}_{\Lambda})$, which implies that $\omega^{\star}_{\Lambda}$ maximises marg_{Λ} in $\Lambda^{>}$. Also, if $\text{marg}_{\Lambda}(w) = \text{marg}_{\Lambda}(\omega^{\star}_{\Lambda})$ then $\|\bar{w}\| = \|\omega^{\star}_{\Lambda}\|$, and thus, $\bar{w} = \omega^{\star}_{\Lambda}$ by uniqueness of $\omega^{\star}_{\Lambda}$. Then $\frac{w}{a_w} = \omega^{\star}_{\Lambda}$ so w is a positive scalar multiple of $\omega^{\star}_{\Lambda}$. Finally, for any $r > 0$, $\text{marg}_{\Lambda}(r\omega^{\star}_{\Lambda}) = \text{marg}_{\Lambda}(\omega^{\star}_{\Lambda})$ so $r\omega^{\star}_{\Lambda}$ maximises marg_{Λ} in $\Lambda^{>}$. □

Results in Section 3.1

The following lemma and proposition are used to prove Proposition 3.

Lemma 35. *Consider any finite $\Lambda \subseteq \mathbb{R}^n$, any $\mathbf{t} \in \mathbb{R}^m_+$, any $r \in \mathbb{R}_+$, and any $v \in \mathbb{R}^n$. If $\mathbf{t}' = \frac{\mathbf{t}}{r}$ then the following results hold.*

(i) *$v \in \Lambda^{\geq}_{\mathbf{t}}$ if and only if rv is in $\Lambda^{\geq}_{\mathbf{t}'}$.*
(ii) *$\omega^{\star}_{\Lambda_{\mathbf{t}'}} = r\omega^{\star}_{\Lambda_{\mathbf{t}}}$; i.e., v has the minimum norm in $\Lambda^{\geq}_{\mathbf{t}}$ if and only if rv has the minimum norm in $\Lambda^{\geq}_{\mathbf{t}'}$.*

Proof. (i): $v \in \Lambda^{\geq}_{\mathbf{t}}$ if and only if for all $i \in I$, $v \cdot (\mathbf{t}_i\lambda_i) \geq 1$, which is if and only if for all $i \in I$, $(\frac{\mathbf{t}_i}{r} rv) \cdot \lambda_i \geq 1$, which holds if and only if for all $i \in I$, $rv \cdot (\mathbf{t}'_i\lambda_i) \geq 1$, which is iff $rv \in \Lambda^{\geq}_{\mathbf{t}'}$.

(ii): v has the minimum norm in $\Lambda^{\geq}_{\mathbf{t}}$ if and only if $v \in \Lambda^{\geq}_{\mathbf{t}}$ and for all $u \in \Lambda^{\geq}_{\mathbf{t}}$, $\|u\| \geq \|v\|$. Part (i) tells us that $v \in \Lambda^{\geq}_{\mathbf{t}}$ $\iff$ $rv \in \Lambda^{\geq}_{\mathbf{t}'}$. Now, for all $u \in \Lambda^{\geq}_{\mathbf{t}}$, $\|u\| \geq \|v\|$ holds if and only if for all $u \in \Lambda^{\geq}_{\mathbf{t}}$, $\|ru\| \geq \|rv\|$ which, from (i), is if and only if for all $ru \in \Lambda^{\geq}_{\mathbf{t}'}$, $\|ru\| \geq \|rv\|$, i.e., for all $u' \in \Lambda^{\geq}_{\mathbf{t}'}$, $\|u'\| \geq \|rv\|$. Thus, v has the minimum norm in $\Lambda^{\geq}_{\mathbf{t}}$ if and only if $v \in \Lambda^{\geq}_{\mathbf{t}}$ and for all $u' \in \Lambda^{\geq}_{\mathbf{t}'}$, $\|u'\| \geq \|rv\|$. This holds if and only if rv has the minimum norm in $\Lambda^{\geq}_{\mathbf{t}'}$. □

To illustrate, consider $\mathbf{t} = (3, 1, 5)$ and $r = 5$. So, $\mathbf{t}' = (3/5, 1/5, 1)$. We know from Example 2 that $\omega^{\star}_{\Lambda_{\mathbf{t}'}} = (1, 2)$ and $\Lambda^{\geq}_{\mathbf{t}'}$ is the intersection of $6x + 3y \geq 5$ and $x + 2y \geq 5$ (i.e., all the shaded region in Fig. 2(a)). It can be shown similarly that $\Lambda^{\geq}_{\mathbf{t}}$ is the intersection of $6x + 3y \geq 1$ and $x + 2y \geq 1$, leading to $\omega^{\star}_{\Lambda_{\mathbf{t}}} = (1/5, 2/5) = \frac{1}{r}\omega^{\star}_{\Lambda'_{\mathbf{t}}}$.

Proposition 36. *Consider any finite consistent set of preference inputs $\Lambda \subseteq \mathbb{R}^n$ any $\mathbf{t} \in \mathbb{R}^m_+$, and any $r \in \mathbb{R}_+$. Then, if $\mathbf{t}' = \frac{\mathbf{t}}{r}$ then $\succcurlyeq^{mm}_{\Lambda_{\mathbf{t}'}}$ is equal to $\succcurlyeq^{mm}_{\Lambda_{\mathbf{t}}}$; i.e., for any α and $\beta \in \mathbb{R}^n$, $\alpha \succcurlyeq^{mm}_{\Lambda_{\mathbf{t}'}} \beta$ if and only if $\alpha \succcurlyeq^{mm}_{\Lambda_{\mathbf{t}}} \beta$.*

Proof. $\alpha \succcurlyeq^{mm}_{\Lambda_{\mathbf{t}'}} \beta$ if and only if $\omega^{\star}_{\Lambda_{\mathbf{t}'}} \cdot \alpha \geq \omega^{\star}_{\Lambda_{\mathbf{t}'}} \cdot \beta$ which, by Lemma 35(ii), holds if and only if $r\omega^{\star}_{\Lambda_{\mathbf{t}}} \cdot \alpha \geq r\omega^{\star}_{\Lambda_{\mathbf{t}}} \cdot \beta$ which is clearly if and only if $\alpha \succcurlyeq^{mm}_{\Lambda_{\mathbf{t}}} \beta$. □

Proposition 3. *Consider any finite consistent set of preference inputs $\Lambda \subseteq \mathbb{R}^n$ and any $\alpha, \beta \in \mathbb{R}^n$. Then, $\alpha \succcurlyeq^{\text{I}}_{\Lambda} \beta$ if and only if for all $\mathbf{t} \in (0, 1]^m$, $\alpha \succcurlyeq^{mm}_{\Lambda_{\mathbf{t}}} \beta$.*

Proof. $\alpha \succcurlyeq^{\text{I}}_{\Lambda} \beta$ iff, by definition of $\succcurlyeq^{\text{I}}_{\Lambda}$, for all $\mathbf{t} \in \mathbb{R}^m_+$, $\alpha \succcurlyeq^{mm}_{\Lambda_{\mathbf{t}}} \beta$. This, by assigning $r = \max_{i \in I} \mathbf{t}_i$ in Proposition 36, holds if and only if for all $\mathbf{t} \in (0, 1]^m$, $\alpha \succcurlyeq^{mm}_{\Lambda_{\mathbf{t}}} \beta$. □

Results in Section 3.2

The proof of Proposition 5 uses the three lemmas below.

Lemma 37. *Consider any finite $\Lambda \subseteq \mathbb{R}^n$, and any $\mathbf{t} \in (0, 1]^m$. Then, for any $u \in \Lambda^{\geq}_{\mathbf{t}}$ we have $\Lambda^* + \{u\} \subseteq \Lambda^{\geq}_{\mathbf{t}}$.*

To illustrate, consider $\mathbf{t} = (3/5, 1/5, 1)$ as in Example 2, and choose $u = (1, 2)$. Then, $\Lambda^* + \{u\}$ is the darkly shaded region in Fig. 2(a). We can see in the figure that $\Lambda^* + \{u\} \subseteq \Lambda^{\geq}_{\mathbf{t}}$ where $\Lambda^{\geq}_{\mathbf{t}}$ is all the shaded region.

Proof. For any $u \in \mathbb{R}^n$ and any $v \in \Lambda^* + \{u\}$, we have, by the definition of Λ^*, $\forall i \in I, (v - u) \cdot \lambda_i \geq 0$, which means that $v \cdot \lambda_i \geq u \cdot \lambda_i$. Also, since it is assumed that $u \in \Lambda^{\geq}_{\mathbf{t}}$, we have $\forall i \in I, u \cdot \lambda_i \geq 1/\mathbf{t}_i$. Thus, $\forall i \in I, v \cdot \lambda_i \geq 1/\mathbf{t}_i$, and so, $v \in \Lambda^{\geq}_{\mathbf{t}}$. □

Lemma 38. *Consider any finite $\Lambda \subseteq \mathbb{R}^n$, and any $u \in \Lambda^{\geq}$. Then, there exists $\mathbf{t} \in (0, 1]^m$ such that $\Lambda^{\geq}_{\mathbf{t}} = \Lambda^* + \{u\}$.*

To illustrate, consider $u = (1, 2)$ and $\mathbf{t} = (1/4, 1/5, 1/3)$. Then, $\Lambda^{\geq}_{\mathbf{t}} = \{(x, y) : \frac{2}{4}x + \frac{1}{4}y \geq 1, \frac{1}{5}x + \frac{2}{5}y \geq 1, \frac{1}{3}x + \frac{1}{3}y \geq 1\}$ which is equal to $\Lambda^* + \{u\}$, the darkly shaded region in Fig. 2(a).

Proof. $u \in \Lambda^{\geq}$ means that for all $i \in I$, $u \cdot \lambda_i \geq 1$, which implies that $0 < \frac{1}{u \cdot \lambda_i} \leq 1$. For all $i \in I$, let $\mathbf{t}_i = \frac{1}{u \cdot \lambda_i}$, and so $\mathbf{t} \in (0, 1]^m$. By definition, $w \in \Lambda^{\geq}_{\mathbf{t}}$ if and only if for all $i \in I$, $w \cdot \mathbf{t}_i \lambda_i \geq 1$. Now, $w \cdot \mathbf{t}_i \lambda_i \geq 1$ holds if and only if $w \cdot \lambda_i \geq u \cdot \lambda_i$, which is if and only if $(w - u) \cdot \lambda_i \geq 0$. Thus, $\Lambda^{\geq}_{\mathbf{t}} = \{w \in \mathbb{R}^n : \forall i \in I, (w - u) \cdot \lambda_i \geq 0\}$, which equals $\Lambda^* + \{u\}$. □

Lemma 39. *Consider any finite $\Lambda \subseteq \mathbb{R}^n$, and any $\mathbf{t} \in (0, 1]^m$. Then, $\Lambda^{\geq}_{\mathbf{t}} \subseteq \Lambda^{\geq}$.*

Proof. Consider any $u \in \Lambda^{\geq}_{\mathbf{t}}$. Then for all $i \in I$, $u \cdot \lambda_i \geq \frac{1}{\mathbf{t}_i}$. Since each $\mathbf{t}_i$ is in $(0, 1]$ we have for all $i \in I$, $u \cdot \lambda_i \geq 1$, and thus, $u \in \Lambda^{\geq}$. □

Proposition 5. *Consider a finite consistent set of preference inputs $\Lambda \subseteq \mathbb{R}^n$ and any $u \in \mathbb{R}^n$. Then, $u \in \mathrm{SI}(\Lambda)$ if and only if $u \in \Lambda^{\geq}$ and u has minimum norm in $\Lambda^* + \{u\}$. Thus, in particular, $\mathrm{SI}(\Lambda) \subseteq \Lambda^{\geq}$.*

Proof. $\Rightarrow$: $u \in \mathrm{SI}(\Lambda)$ means that there exists $\mathbf{t} \in (0, 1]^m$ such that $u \in \Lambda^{\geq}_{\mathbf{t}}$ and u has the minimum norm in $\Lambda^{\geq}_{\mathbf{t}}$, which, since $\Lambda^{\geq}_{\mathbf{t}} \subseteq \Lambda^{\geq}$ by Lemma 39, implies that $u \in \Lambda^{\geq}$. Now, u also has the minimum norm in $\Lambda^* + \{u\}$ because firstly, $\Lambda^* + \{u\} \subseteq \Lambda^{\geq}_{\mathbf{t}}$ from Lemma 37, and secondly, u is clearly in $\Lambda^* + \{u\}$ since $\vec{\mathbf{0}} \in \Lambda^*$.

$\Leftarrow$: Assume now that $u \in \Lambda^{\geq}$ and u has the minimum norm in $\Lambda^* + \{u\}$. By Lemma 38, there exists $\mathbf{t} \in (0, 1]^m$ such that u has the minimum norm in $\Lambda^{\geq}_{\mathbf{t}} (= \Lambda^* + \{u\})$, and clearly $u \in \Lambda^{\geq}_{\mathbf{t}}$. Thus, $u \in \mathrm{SI}(\Lambda)$. □

The following lemma is used in the proof of Proposition 6; it states a basic property of a minimal norm element in a convex set.

Lemma 40. *Consider any $u \in G$ where $G \subseteq \mathbb{R}^n$ is a convex set. Then, u has the minimum norm in G if and only if for all $v \in G$, $u \cdot (v - u) \geq 0$.*

Proof. $\Rightarrow$: Firstly, for the case when $v = u$, the result is easily obtained because $u \cdot (v - u) = 0$. Now, consider any $v \in G - \{u\}$. We define $v_\delta = \delta v + (1 - \delta)u$ for each $\delta \in (0, 1]$. It is clear that $v_\delta \in G$ because v and u both are in the convex set G, and since u has the minimum norm in G, for all $\delta \in (0, 1]$ we have that $\|v_\delta\| \geq \|u\|$. Now, assume that $u \cdot (v - u) < 0$. We show that this assumption leads to $\|v_\delta\| < \|u\|$ for some $\delta \in (0, 1]$, which will prove the first part by contradiction. To do this, we rewrite $\|v_\delta\|^2 - \|u\|^2$ as follows:

$$\begin{aligned}\|v_\delta\|^2 - \|u\|^2 &= \|\delta(v - u) + u)\|^2 - \|u\|^2 \\ &= (\delta(v - u) + u)) \cdot (\delta(v - u) + u)) - u \cdot u,\end{aligned}$$

which equals $\delta^2(v - u) \cdot (v - u) + 2\delta u \cdot (v - u)$, i.e., $\delta(\delta\|v - u\|^2 + 2u \cdot (v - u))$. Now, since $u \cdot (v - u) < 0$, for sufficiently small δ, $\|v_\delta\|^2 - \|u\|^2 < 0$, and thus $\|v_\delta\| < \|u\|$.

$\Leftarrow$: Consider any $v \in G - \{u\}$. Since $u \neq v$, $\|v - u\|^2 > 0$, which implies that $(v - u) \cdot (v - u) > 0$, and thus, $\|v\|^2 + \|u\|^2 > 2v \cdot u$. Also, $u \cdot (v - u) \geq 0$ leads to $v \cdot u \geq \|u\|^2$. Hence, $\|v\|^2 + \|u\|^2 > 2\|u\|^2$, and thus, $\|v\| > \|u\|$, showing that u has minimum norm in G. □

Proposition 6. *Consider any finite consistent set of preference inputs $\Lambda \subseteq \mathbb{R}^n$ and any $u \in \mathbb{R}^n$. Then, u has minimum norm in $\Lambda^* + \{u\}$ if and only if $u \in co(\Lambda)$.*

Proof. Clearly, $\Lambda^* + \{u\}$ is a convex set. Lemma 40 implies that u has minimum norm in $\Lambda^* + \{u\}$ if and only if for all $v \in \Lambda^* + \{u\}$, $u \cdot (v - u) \geq 0$. By writing $y = v - u$, this is if and only if for all $y \in \Lambda^*$, $u \cdot y \geq 0$, which holds if and only if for all $y \in \Lambda^*$, $y \in \{u\}^*$. Thus, u has minimum norm in $\Lambda^* + \{u\}$ if and only if $\Lambda^* \subseteq \{u\}^*$. Lemma 33 then implies the result. □

Results in Section 4.1

The following lemma is used to prove the equivalence in Proposition 11.

Lemma 41. *Consider any $v \in \mathbb{R}^n$ and any $\tau \in \mathbb{R}^n_+$. Then, $v \in \Lambda^{\geq}$ if and only if $v \odot \tau^{-1} \in (\Lambda \odot \tau)^{\geq}$. Also, $w = v$ minimises $\|w \odot \tau^{-1}\|$ over $w \in \Lambda^{\geq}$ if and only if $v = \tau \odot \omega^{\star}_{\Lambda \odot \tau}$.*

For example, we have seen in Example 3 that for $\tau = (2, 3)$, $\omega^{\star}_{\Lambda\odot\tau} = (2/13, 3/13)$. This lemma implies that for $v = \tau \odot \omega^{\star}_{\Lambda\odot\tau} = (4/13, 9/13)$, there is no $u \in \Lambda^{\geq}$ such that $\|u \odot \tau^{-1}\| < \|v \odot \tau^{-1}\|$, i.e., v will be rescale-optimal in $\Lambda^{\geq}$.

Proof. The first part follows easily from the definitions. Regarding the second part, by definition of $\omega^{\star}_{\Lambda\odot\tau}$, we have that $v = \tau \odot \omega^{\star}_{\Lambda\odot\tau}$ if and only if $v \odot \tau^{-1}$ has minimum norm in $(\Lambda \odot \tau)^{\geq}$, which is if and only if $w' = v \odot \tau^{-1}$ minimises $\|w'\|$ over $w' \in \{w' : \forall\lambda \in \Lambda, w' \cdot \lambda \odot \tau \geq 1\}$. By substituting w' with $w \odot \tau^{-1}$ this holds if and only if $v = w$ minimises $\|w \odot \tau^{-1}\|$ over $w \in \Lambda^{\geq}$. □

Proposition 11. *Consider any finite consistent set of preference inputs $\Lambda \subseteq \mathbb{R}^n$. Then, $\mathrm{SF}(\Lambda)$ is equal to the set of all rescale-optimal elements of $\Lambda^{\geq}$. Thus, for $\alpha, \beta \in \mathbb{R}^n$, $\alpha \succcurlyeq^{\mathrm{F}}_{\Lambda} \beta$ if and only if $w \cdot (\alpha - \beta) \geq 0$ for every rescale-optimal element w in $\Lambda^{\geq}$.*

Proof. Consider any $u \in \mathbb{R}^n$. Then, u is rescale-optimal in $\Lambda^{\geq}$ if and only if there exists $\tau \in \mathbb{R}^n_+$ such that $u = w$ minimises $\|w \odot \tau\|$ over $w \in \Lambda^{\geq}$, which, by Lemma 41, is if and only if there exists $\tau \in \mathbb{R}^n_+$ such that $u = \tau^{-1} \odot \omega^{\star}_{\Lambda\odot\tau^{-1}}$, which is, from the definition of $\mathrm{SF}(\Lambda)$, if and only if $u \in \mathrm{SF}(\Lambda)$. □

Results in Section 4.3

The proof of Theorem 13 uses a triple of lemmas.

Lemma 42. *Let G be a convex and (topologically) closed subset of $\mathbb{R}^n$. For each strictly positive vector $\tau \in \mathbb{R}^n_+$, there exists a unique $w \in G$ with minimum value of $\|w \odot \tau\|$.*

Proof. It is a standard result (for a proof see e.g., Proposition 4 of [49]) that there is a unique element in a convex closed set with minimum norm. Consider any $\tau \in \mathbb{R}^n_+$. Now, $G \odot \tau = \{w \odot \tau : w \in G\}$ is convex and closed so there exists a unique element $w \odot \tau \in G \odot \tau$ with minimum value of $\|w \odot \tau\|$, so there is a unique $w \in G$ with minimum value of $\|w \odot \tau\|$. □

Lemma 43. *Let $u, v \in \mathbb{R}^n$. There exists $k \in \{1, \ldots, n\}$ such that $|u(k)| < |v(k)|$ if and only if there exists $\tau \in \mathbb{R}^n_+$ such that $\|u \odot \tau\| < \|v \odot \tau\|$. Thus, for all $j \in \{1, \ldots, n\}$, $|u(j)| \geq |v(j)|$ if and only if for all $\tau \in \mathbb{R}^n_+$, $\|u \odot \tau\| \geq \|v \odot \tau\|$.*

Proof. $\Rightarrow$: Assume first that there exists $k \in \{1, \ldots, n\}$ such that $|u(k)| < |v(k)|$. For $\epsilon > 0$, define $\tau_\epsilon \in \mathbb{R}^n_+$ by $\tau_\epsilon(k) = 1 + \epsilon$, and, for $j \neq k$, $\tau_\epsilon(j) = \epsilon$. Then $u \odot \tau_\epsilon = \epsilon u + u(k)e_k$, where e_k is the unit vector in the kth direction, which leads to $\|u \odot \tau_\epsilon\|^2 = (u \odot \tau_\epsilon) \cdot (u \odot \tau_\epsilon)$ equalling $\epsilon^2 u \cdot u + (1 + 2\epsilon)u(k)^2$. Similarly, $\|v \odot \tau_\epsilon\|^2 = \epsilon^2 v \cdot v + (1 + 2\epsilon)v(k)^2$. Since $u(k)^2 < v(k)^2$, for sufficiently small $\epsilon > 0$, we will have $\|u \odot \tau_\epsilon\|^2 < \|v \odot \tau_\epsilon\|^2$.

$\Leftarrow$: Now assume that there exists $\tau \in \mathbb{R}^n_+$ such that $\|u \odot \tau\| < \|v \odot \tau\|$. Then for some $k \in \{1, \ldots, n\}$, $|(u \odot \tau)(k)| < |(v \odot \tau)(k)|$, i.e., $|u(k)\tau(k)| < |v(k)\tau(k)|$, which implies $|u(k)| < |v(k)|$, since $\tau(k)$ is non-zero. □

Lemma 15. *Let G be a convex subset of $\mathbb{R}^n$, and let j be any element of $\{1, \ldots, n\}$. Then either (i) there exists $w \in G$ such that $w(j) = 0$; or (ii) for all $w \in G$, $w(j) > 0$; or (iii) for all $w \in G$, $w(j) < 0$.*

Proof. To prove a contradiction, suppose that neither (i), (ii) nor (iii) hold for j, so for all $w \in G$, $w(j) \neq 0$, and there exists $u, v \in G$ such that $u(j) > 0$ and $v(j) < 0$. Let $\delta = \frac{u(j)}{u(j)-v(j)}$, so that $1 - \delta = \frac{-v(j)}{u(j)-v(j)}$. Let $v_\delta = \delta v + (1 - \delta)u$, which is in G because G is convex and $\delta \in (0, 1)$. Then, $v_\delta(j) = 0$, which shows that (i) holds for j, contradicting the earlier assumption. □

Theorem 13. *Let G be a convex and closed subset of $\mathbb{R}^n$, and let u be an element of G. Then the following conditions are equivalent.*

(i) *u is uniquely rescale-optimal in G, i.e., u is the unique element of G that is rescale-optimal;*
(ii) *for all $v \in G$, for all $j \in \{1, \ldots, n\}$, $|v(j)| \geq |u(j)|$;*
(iii) *u pointwise dominates every element in $G - \{u\}$.*

Proof. First suppose (i), that u is uniquely rescale-optimal in G, so that, for all $\tau \in \mathbb{R}^n_+$, and for all $v \in G - \{u\}$, $\|u \odot \tau\| \leq \|v \odot \tau\|$; thus, by Lemma 43, for all $j \in \{1, \ldots, n\}$, $|u(j)| \leq |v(j)|$, showing that (ii) holds. The converse follows easily: if (ii) holds then for all $\tau \in \mathbb{R}^n_+$, for all $v \in G$, $\|u \odot \tau\| \leq \|v \odot \tau\|$, which by Lemma 42, leads to for all $\tau \in \mathbb{R}^n_+$, for all $v \in G - \{u\}$, $\|u \odot \tau\| < \|v \odot \tau\|$, and thus proving (i).

We will next show that (ii) implies (iii). Consider any $j \in \{1, \ldots, n\}$. We must show that for each $v \in G - \{u\}$, either $0 \leq u(j) \leq v(j)$ or $0 \geq u(j) \geq v(j)$ hold. This holds trivially if $u(j) = 0$, so suppose $u(j) \neq 0$. Then, (ii) implies that, for all $v \in G$, $v(j) \neq 0$, and thus, by Lemma 15, either, for all $v \in G$, $v(j) > 0$, or for all $v \in G$, $v(j) < 0$. (ii) then implies that for all $v \in G$, either $0 \leq u(j) \leq v(j)$ or $0 \geq u(j) \geq v(j)$, proving (iii). It immediately follows that (iii) implies (ii), completing the proof of equivalence of (i), (ii) and (iii). □

Results in Section 4.4

The definitions easily imply the following lemma, which relates pointwise dominance and zm-pointwise dominance.

Lemma 44. *Consider any $u, v \in \mathbb{R}^n$. If v pointwise dominates u then v zm-pointwise dominates u.*

Now suppose that $u \in G \subseteq \mathbb{R}^n$. If u is zm-pointwise undominated in G then u is pointwise undominated in G. In addition, the converse holds if none of the components of u is zero.

Proof. Suppose that v pointwise dominates u. Then, $u \neq v$ and for all $j \in \{1, \ldots, n\}$, either (i) $0 \leq v(j) \leq u(j)$, or (ii) $0 \geq v(j) \geq u(j)$. So, for some $k \in \{1, \ldots, n\}$, $u(k) \neq v(k)$, which implies that $u(k) \neq 0$. This implies that v zm-pointwise dominates u. The other two parts follow immediately from the definitions. □

Lemma 45 below characterises the zm-pointwise undominated elements in a convex set.

Lemma 45. *Consider any convex set $G \in \mathbb{R}^n$. Then, u is zm-pointwise undominated in convex G if and only if for all $v \in G$, either*

(i) *$v(j) = u(j)$ for all $j \in \{1, \ldots, n\}$ such that $u(j) \neq 0$; or*
(ii) *there exists $k \in \{1, \ldots, n\}$ such that either $0 < u(k) < v(k)$ or $0 > u(k) > v(k)$.*

Proof. First, let us suppose that u is not zm-pointwise undominated in G. We will show that there exists $v \in G$ such that neither condition (i) nor condition (ii) hold for v. Since u is not zm-pointwise undominated in G, there exists $v \in G$ that zm-pointwise dominates u. By definition, there exists $j \in \{1, \ldots, n\}$ such that $v(j) \neq u(j) \neq 0$, and thus, condition (i) does not hold for v; also for all $k \in \{1, \ldots, n\}$ with $u(k) \neq 0$, either $0 \leq v(k) \leq u(k)$ or $0 \geq v(k) \geq u(k)$, which means that condition (ii) in this lemma does not hold for v.

Conversely, suppose that it is not the case that for all $v \in G$, either (i) $v(j) = u(j)$ for all $j \in \{1, \ldots, n\}$ such that $u(j) \neq 0$; or (ii) there exists $k \in \{1, \ldots, n\}$ such that either $0 < u(k) < v(k)$ or $0 > u(k) > v(k)$. Then, there exists $v \in G$ such that (i) there exists $k \in \{1, \ldots, n\}$ such that $u(k) \neq 0$ and $u(k) \neq v(k)$; and (ii) for all $j \in \{1, \ldots, n\}$, if $u(j) > 0$ then $v(j) \leq u(j)$; and if $u(j) < 0$ then $v(j) \geq u(j)$. Thus, there exists $v \in G$ such that (i) there exists $k \in N_u$ such that $u(k) \neq v(k)$; and (ii) for all $j \in N_u$, if $u(j) > 0$ then $v(j) \leq u(j)$; and if $u(j) < 0$ then $v(j) \geq u(j)$ (where $N_u = \{j \in \{1, \ldots, n\} : u(j) \neq 0\}$, as in Definition 9). For $\delta \in (0, 1]$ let $v_\delta = \delta v + (1 - \delta)u$, which is in G. Then there exists $\delta \in (0, 1]$ such that (i) there exists $k \in N_u$ such that $u(k) \neq v_\delta(k)$; and (ii) for all $j \in N_u$, if $u(j) > 0$ then $0 < v_\delta(j) \leq u(j)$; and if $u(j) < 0$ then $0 > v_\delta(j) \geq u(j)$. Thus, v_δ zm-pointwise dominates u showing that u is not pointwise undominated in G. □

Lemma 46 is used in the proof of Proposition 16.

Lemma 46. *Let $u, v \in \mathbb{R}^n$, with $u \neq v$, and let $\tau \in \mathbb{R}^n_+$. For $\delta \in (0, 1]$ let $v_\delta = \delta v + (1 - \delta)u$. Then the following hold:*

(i) *For any $\delta \in (0, 1]$, $\|v_\delta \odot \tau\|^2 - \|u \odot \tau\|^2 = \delta^2 \|(v - u) \odot \tau\|^2 + 2\delta(\tau \odot \tau \odot u) \cdot (v - u)$.*
(ii) *$(\tau \odot \tau \odot u) \cdot (v - u) \geq 0$ if and only if for all $\delta \in (0, 1]$, $\|v_\delta \odot \tau\| > \|u \odot \tau\|$.*
(iii) *There exists $\tau \in \mathbb{R}^n_+$ such that $(\tau \odot \tau \odot u) \cdot (v - u) \geq 0$ if and only if either (a) $v(j) = u(j)$ for all $j \in \{1, \ldots, n\}$ such that $u(j) \neq 0$; or (b) there exists $k \in \{1, \ldots, n\}$ such that either $0 < u(k) < v(k)$ or $0 > u(k) > v(k)$.*

Proof. (i): Using $v_\delta = u + \delta(v - u)$, we have that $v_\delta \odot \tau = (u \odot \tau) + \delta(v - u) \odot \tau$. Then, $\|v_\delta \odot \tau\|^2 = (v_\delta \odot \tau) \cdot (v_\delta \odot \tau) = (u \odot \tau) \cdot (u \odot \tau) + \delta^2 \|(v - u) \odot \tau\|^2 + 2\delta(u \odot \tau) \cdot ((v - u) \odot \tau)$, which leads to the result.

(ii): If $(\tau \odot \tau \odot u) \cdot (v - u) \geq 0$ then (i) immediately implies that for all $\delta \in (0, 1]$, $\|v_\delta \odot \tau\| > \|u \odot \tau\|$, since $\|(v - u) \odot \tau\|$ is non-zero (because $u \neq v$). Conversely, suppose that $(\tau \odot \tau \odot u) \cdot (v - u) < 0$. Choosing δ such that $\delta \|(v - u) \odot \tau\|^2 \leq -2(\tau \odot \tau \odot u) \cdot (v - u)$ gives, using (i), that $\|v_\delta \odot \tau\| \leq \|u \odot \tau\|$, proving (ii).

(iii), $\Rightarrow$: Suppose that there exists $\tau \in \mathbb{R}^n_+$ such that $(\tau \odot \tau \odot u) \cdot (v - u) \geq 0$, and it is not the case that there exists $k \in \{1, \ldots, n\}$ such that either $0 < u(k) < v(k)$ or $0 > u(k) > v(k)$. Then, we can see that for each $j \in \{1, \ldots, n\}$, $(\tau \odot \tau \odot u)(j)(v - u)(j) = \tau(j)^2 u(j)(v(j) - u(j)) \leq 0$ (since it clearly holds if $u(j) = 0$; if $u(j) > 0$ then $v(j) \leq u(j)$ so it also holds; if $u(j) < 0$ then $v(j) \geq u(j)$ so it holds then too). The sum (over each j) of these n terms is at least zero, since it is equal to $(\tau \odot \tau \odot u) \cdot (v - u)$ and thus, for each $j \in \{1, \ldots, n\}$, $(\tau \odot \tau \odot u)(j)(v - u)(j) = 0$. This implies that $v(j) = u(j)$ for all $j \in \{1, \ldots, n\}$ such that $u(j) \neq 0$.

$\Leftarrow$: If (a) $v(j) = u(j)$ for all $j \in \{1, \ldots, n\}$ such that $u(j) \neq 0$ then $(\tau \odot \tau \odot u) \cdot (v - u) \geq 0$. Now, assume that (b) holds, i.e., there exists $k \in \{1, \ldots, n\}$ such that either $0 < u(k) < v(k)$ or $0 > u(k) > v(k)$. For $\epsilon > 0$, define τ_ϵ by $\tau_\epsilon(k) = \sqrt{1 + \epsilon}$, and $\tau_\epsilon(j) = \sqrt{\epsilon}$ for all $j \neq k$. Then, $\tau_\epsilon \odot \tau_\epsilon \odot u = u(k)e_k + \epsilon u$, where e_k is the unit vector in the kth direction, so $(\tau_\epsilon \odot \tau_\epsilon \odot u) \cdot (v - u) = u(k)(v(k) - u(k)) + \epsilon u \cdot (v - u)$, which is greater than zero for sufficiently small ϵ, since $u(k)(v(k) - u(k)) > 0$. □

Proposition 16. *Let u be an element of convex $G \subseteq \mathbb{R}^n$. Then:*

(i) *u is rescale-optimal in G if and only if there exists $\tau \in \mathbb{R}^n_+$ such that for all $v \in G$, $(\tau \odot \tau \odot u) \cdot (v - u) \geq 0$ (i.e., $(\tau \odot u) \cdot (\tau \odot (v - u)) \geq 0$).*
(ii) *u is zm-pointwise undominated in G if and only if for all $v \in G$, there exists $\tau \in \mathbb{R}^n_+$ such that $(\tau \odot \tau \odot u) \cdot (v - u) \geq 0$.*
(iii) *If u is rescale-optimal in G then u is zm-pointwise undominated in G.*

Proof. (i): Using Lemma 42, u is rescale-optimal in G if and only if there exists $\tau \in \mathbb{R}^n_+$ such that for all $v \in G - \{u\}$, $\|v \odot \tau\| > \|u \odot \tau\|$, which, since G is convex, is if and only if, there exists $\tau \in \mathbb{R}^n_+$ such that for all $v \in G - \{u\}$ and for all $\delta \in (0, 1]$, $\|v_\delta \odot \tau\| > \|u \odot \tau\|$, where $v_\delta = \delta v + (1 - \delta)u$. By Lemma 46(ii), this is if and only if there exists $\tau \in \mathbb{R}^n_+$ such that for all $v \in G - \{u\}$, $(\tau \odot \tau \odot u) \cdot (v - u) \geq 0$, which holds if and only if there exists $\tau \in \mathbb{R}^n_+$ such that for all $v \in G$, $(\tau \odot \tau \odot u) \cdot (v - u) \geq 0$.

(ii) By Lemma 45 and Lemma 46(iii), u is zm-pointwise undominated in G if and only if for all $v \in G - \{u\}$, there exists $\tau \in \mathbb{R}^n_+$ such that $(\tau \odot \tau \odot u) \cdot (v - u) \geq 0$, from which (ii) follows.

(iii) follows immediately from (i) and (ii). □

Results in Section 4.5

The following lemmas are used in the proof of Theorem 18.

Lemma 47. *Consider any $u \in G_I$ and any $v \in G_{J_u}$. Then there exists $\delta' \in (0, 1)$ such that for all δ with $0 < \delta \leq \delta'$, $\delta v + (1 - \delta)u \in G_I$.*

To illustrate, consider $u = (1, 0)$ and $v = (-1, 2)$ which is in G_{J_u}. We can see in Fig. 1(b) that the line segment from u to $(0, 1)$ is in G_I but beyond that from $(0, 1)$ to v is not. That means that choosing $\delta' = 1/2$ works for this case (because $\frac{1}{2}v + (1 - \frac{1}{2})u = (0, 1)$); i.e., for all δ with $0 < \delta \leq 1/2$, $\delta v + (1 - \delta)u \in G_I$.

Proof. Let $x = v - u$, and, for all $\delta \in (0, 1)$, let $v_\delta = u + \delta x = \delta v + (1 - \delta)u$. Since $u, v \in G_{J_u}$, we have $v_\delta \in G_{J_u}$ for all $\delta \in (0, 1)$. We will next show that for all sufficiently small δ, $v_\delta \in G_I$, i.e., that for all $i \in I$, $v_\delta \cdot \lambda_i \geq a_i$. Since, $v_\delta \in G_{J_u}$, this holds for all $i \in J_u$. Now, consider any $i \in I - J_u$. By definition of J_u we have $u \cdot \lambda_i > a_i$. This implies that there exists $\delta_i > 0$ with for all δ with $0 < \delta \leq \delta_i$, $(u \cdot \lambda_i) + \delta(x \cdot \lambda_i) > a_i$, and thus $v_\delta \cdot \lambda_i > a_i$. Let us choose $\delta' = \min\{\delta_i : i \in I - J_u\}$. Then for all δ with $0 < \delta \leq \delta'$, and for all $i \in I - J_u$, $v_\delta \cdot \lambda_i > a_i$, so for all $i \in I$, $v_\delta \cdot \lambda_i \geq a_i$, which shows that $v_\delta \in G_I$. □

Lemma 48. *Consider non-zero $u \in G_I$ (as defined above). Then u is zm-pointwise undominated in G_I if and only if u is zm-pointwise undominated in G_{J_u}.*

Proof. $\Rightarrow$: Suppose that u is zm-pointwise undominated in G_I. Consider any $v \in G_{J_u}$. By Lemma 47, there exists $\delta \in (0, 1)$ such that $v_\delta \in G_I$, where $v_\delta = \delta v + (1 - \delta)u = u + \delta(v - u)$. Proposition 16(ii) implies that there exists $\tau \in \mathbb{R}^n_+$ such that for all $w \in G_I$, $(\tau \odot \tau \odot u) \cdot (w - u) \geq 0$. In particular, $(\tau \odot \tau \odot u) \cdot (v_\delta - u) \geq 0$, i.e., $(\tau \odot \tau \odot u) \cdot (\delta(v - u)) \geq 0$, which implies that $(\tau \odot \tau \odot u) \cdot (v - u) \geq 0$. Note that δ does not depend on the choice of v. Thus, there exists $\tau \in \mathbb{R}^n_+$ such that for all $v \in G_{J_u}$, $(\tau \odot \tau \odot u) \cdot (v - u) \geq 0$. Applying Proposition 16(ii) again gives that u is zm-pointwise undominated in G_{J_u}.

$\Leftarrow$: This is immediate because $G_I \subseteq G_{J_u}$. □

Lemma 19. *$G_{J_u} + \{-u\}$ is equal to $\{\lambda_i : i \in J_u\}^*$.*

$G_{J_u} + \{-u\}$ means translating G_{J_u} to move u to the origin. So, continuing the example, for $u = (1, 0)$, $G_{J_u} + \{-u\} = \{w \in \mathbb{R}^2 : w \cdot (1, 1) \geq 0, w \cdot (1, 2) \geq 0\} = \{(1, 1), (1, 2)\}^* = \{\lambda_i : i \in J_u\}^*$.

Proof. $v \in G_{J_u} + \{-u\}$ if and only if for all $i \in J_u$, $(v + u) \cdot \lambda_i \geq a_i$, which is if and only if for all $i \in J_u$, $v \cdot \lambda_i \geq 0$ (since, by definition, $u \cdot \lambda_i = a_i$ for all $i \in J_u$), which is if and only if $v \in (\{\lambda_i : i \in J_u\})^*$. □

Lemma 49. *For $u, v \in \mathbb{R}^n$, if u and v agree on signs and $u \neq \vec{\mathbf{0}}$ then $u \cdot v > 0$.*

Proof. Because $u \neq \vec{\mathbf{0}}$, there exists $k \in \{1, \ldots, n\}$ with $u(k) \neq 0$. Then, since u and v agree on signs, $v(k)$ is non-zero and the same sign as $u(k)$, so $u(k)v(k) > 0$. Also, for any $j \in \{1, \ldots, n\}$, $u(j)v(j) \geq 0$, and thus, $u \cdot v > 0$. □

Theorem 18. *Let u be an element of polyhedron $G \subseteq \mathbb{R}^n$. Then, u is rescale-optimal in G if and only if u is zm-pointwise undominated in G.*

Proof. G is a polyhedron, so, by definition, it can be written as $\{w \in \mathbb{R}^n : \forall i \in I, w \cdot \lambda_i \geq a_i\}$. Let $J_u = \{i \in I : \lambda_i \cdot u = a_i\}$, and let $G_{J_u} = \{w \in \mathbb{R}^n : \forall i \in J_u, w \cdot \lambda_i \geq a_i\}$. Proposition 16(iii) implies that if u is rescale-optimal in G then u is zm-pointwise undominated. We next prove the converse.

Assume that u is zm-pointwise undominated in G. Let $C = G_{J_u} + \{-u\}$. By Lemma 19, $C = \{\lambda_i : i \in J_u\}^*$, which is a polyhedral cone (i.e., a polyhedron that is cone), and thus, by the Minkowski-Weyl theorem (see e.g., Theorem 4.18 of [57]), is a finitely generated convex cone, so we can write $C = co(W)$ for some finite set $W = \{w_1, \ldots, w_l\}$.

Let $C' = co(S)$ be the convex cone generated by $S = W \cup S_Z$ where $S_Z = \{e_j, -e_j : j \in Z\}$, and $e_j \in \mathbb{R}^n$ is the unit vector in the jth dimension, and $Z = \{j \in \{1, \ldots, n\} : u(j) = 0\}$. Also, let $T = E_+ \cup E_- \cup R$, where $E_+ = \{-e_j : u(j) > 0\}$, and $E_- = \{e_j : u(j) < 0\}$, and $R = \{-w_i : i \in M\}$, and where $M = \{i \in L : -w_i \notin C'\}$ and $L = \{1, \ldots, l\}$. Let H be the convex hull of T.

We will show that the assumption that u is zm-pointwise undominated implies that C' and H are disjoint. If there exists $h \in C' \cap H$ then h can be written as $w + v_0$ where $w \in C$ and $v_0 \in co(S_Z)$. Also, since $h \in H$, it can be written as $v_+ + v_- + y$, where $v_+ \in co(E_+)$, $v_- \in co(E_-)$ and $y \in co(R)$. (More specifically, for some $q_1, q_2, q_3 \in [0, 1]$ with $q_1 + q_2 + q_3 = 1$ we have $v_+ = q_1 v'_+$ for some v'_+ in the convex hull of E_+, and $v_- = q_2 v'_-$ for some v'_- in the convex hull of E_-, and $y = q_3 z$ for some z in the convex hull of R.) Since $-y \in C$, $w - y \in C$. Let $v = w - y + u = -v_0 + v_+ + v_- + u$. Then $v \in G_{J_u}$, because $v - u = w - y \in C$.

For $j \in \{1, \ldots, n\}$, if $u(j) > 0$ then $v_0(j) = v_-(j) = 0$, so $v(j) = u(j) + v_+(j) \leq u(j)$. Similarly, if $u(j) < 0$ then $v(j) = u(j) + v_-(j) \geq u(j)$. Thus, if $u(j) > 0$ then $v(j) \leq u(j)$; and if $u(j) < 0$ then $v(j) \geq u(j)$. Since, u is zm-pointwise undominated in G, u is zm-pointwise undominated in G_{J_u}, by Lemma 48. Lemma 45 then implies that for all $j \in \{1, \ldots, n\}$, if $u(j) \neq 0$ then $v(j) = u(j)$, and thus $v_+(j) = v_-(j) = 0$, and so, $v_+ = v_- = \vec{\mathbf{0}}$ (since also, if $u(j) = 0$ then $v_+(j) = v_-(j) = 0$, by definition of v_+ and v_-, and of E_+ and E_-). This implies that $w + v_0 = y$ and $y \in H$. Also, since $\vec{\mathbf{0}}$ is neither in the convex hull of E_+ nor E_-, we have $q_1 = q_2 = 0$, and thus $q_3 = 1$, and so, y is in the convex hull of R. By definition of convex hull, we can write y as $\sum_{i \in M} t_i(-w_i)$, with each $t_i \geq 0$, and for some $k \in M$, $t_k > 0$. Then $-t_k w_k = w + \sum_{i \in M, i \neq k} t_i w_i + v_0$. The right-hand-side is in $co(S)$, which equals C', which implies that $-w_k \in C'$, which contradicts $k \in M$. Thus, C' and H are disjoint.

Both C' and H are convex and closed, and H is compact. A strict separating hyperplane theorem (see e.g., Theorem 2.1.5 of [58]) implies that there exists vector $\mu \in \mathbb{R}^n$ and $c \in \mathbb{R}$ such that for all $g \in C'$, $\mu \cdot g > c$ and for all $h \in H$, $\mu \cdot h < c$. Since $\vec{\mathbf{0}} \in C'$, we have $\mu \cdot \vec{\mathbf{0}} > c$, so $c < 0$.

Now, if g and $-g$ are both in C' then $\mu \cdot g = 0$. (Else $\mu \cdot g < 0$ or $\mu \cdot (-g) < 0$; without loss of generality assume $\mu \cdot g < 0$; then there exists $r > 0$ such that $\mu \cdot (rg) = r(\mu \cdot g) < c$, which contradicts $rg \in C'$.) This implies that if $u(j) = 0$ (so $j \in Z$ and $e_j, -e_j \in C'$) then $\mu \cdot e_j = 0$ and thus $\mu(j) = 0$. Also, if $i \in L - M$, then $w_i, -w_i \in C'$, so $\mu \cdot w_i = 0$. For any $i \in M$, we have that $-w_i \in H$, so $\mu \cdot (-w_i) < c < 0$, so $\mu \cdot w_i > 0$. Thus for any $w_i \in W$, $\mu \cdot w_i \geq 0$, and therefore for any $w \in C$, $\mu \cdot w \geq 0$, since w is a positive linear combination of the elements of W.

If $u(j) > 0$, then $-e_j \in H$, so $\mu \cdot e_j > -c > 0$, so $\mu(j) > 0$. Similarly, if $u(j) < 0$ then $\mu(j) < 0$. Thus, μ agrees on signs with u. This, by using Lemma 49, implies that $\mu \cdot u > 0$ (since $u \neq \vec{\mathbf{0}}$), and we let $\mu' = \frac{\mu}{\mu \cdot u}$. So $\mu' \cdot u = 1$, and μ' agrees on signs with μ and then u too.

For any $v \in G$, $v \in G_{J_u}$, and so $v - u$ is in C; we have shown that $\mu' \cdot (v - u) \geq 0$, so $\mu' \cdot v \geq \mu' \cdot u = 1$. Theorem 17 then implies that u is rescale-optimal in G. □

Results in Section 4.6

Lemma 20. *Consider a polyhedron G_I and non-zero $u \in G_I$. Then u is rescale-optimal in G_I if and only if u is rescale-optimal in G_{J_u}.*

This follows from Theorem 18 and Lemma 48, since G_I and G_{J_u} are polyhedra. However, we give a more direct proof here.

Proof. Firstly, since $G_I \subseteq G_{J_u}$, if u is rescale-optimal in G_{J_u} then u is rescale-optimal in G_I (since the same scaling function τ can be used). We will go on to prove the converse; so, let us assume that u is rescale-optimal in G_I. Theorem 17 implies that there exists $\mu \in \mathbb{R}^n$ agreeing on signs with u such that $\mu \cdot u = 1$ and for all $w \in G_I$, $\mu \cdot w \geq 1$. Consider arbitrary $v \in G_{J_u}$; we will show that $\mu \cdot v \geq 1$.

Let $x = v - u$, and, for all $\delta \in (0, 1)$, let $v_\delta = \delta v + (1 - \delta)u = u + \delta x$. By Lemma 47, there exists $\delta \in (0, 1)$ such that $v_\delta \in G_I$. This implies that $\mu \cdot v_\delta \geq 1$, so $\mu \cdot u + \delta \mu \cdot x \geq 1$, and hence $\delta \mu \cdot x \geq 0$ and $\mu \cdot x \geq 0$, and therefore, $\mu \cdot v \geq \mu \cdot u = 1$.

We have shown that for all $v \in G_{J_u}$, $\mu \cdot v \geq 1$; we also have that μ and u agree on signs and $\mu \cdot u = 1$. Using Theorem 17, this implies that u is rescale-optimal in G_{J_u}, as required. □

Results in Section 6

Proposition 27. *Consider any finite subset Λ of $\mathbb{R}^n$, and define $\mu = \sum_{\lambda \in \Lambda} \lambda$ and define Λ_μ to be $\{\lambda \in \Lambda : \lambda \cdot \mu > 0\}$. Assume that $\mu \neq \vec{\mathbf{0}}$. Then the following hold.*

(i) *Λ_μ is non-empty and consistent, i.e., $(\Lambda_\mu)^> \neq \emptyset$.*
(ii) *Let $\omega^\star_{\Lambda_\mu}$ be the solution of the maximum margin approach for Λ_μ, i.e., the minimal norm element in $(\Lambda_\mu)^\geq$. Then $\Lambda_{\omega^\star_{\Lambda_\mu}} = \{\lambda \in \Lambda : \lambda \cdot \omega^\star_{\Lambda_\mu} > 0\}$ is non-empty and consistent, and $\Lambda_\mu \subseteq \Lambda_{\omega^\star_{\Lambda_\mu}} \subseteq \Lambda$.*

Proof. (i): We will use proof by contradiction to show that Λ_μ is non-empty. Suppose the contrary, that $\Lambda_\mu = \emptyset$. Then $\lambda \cdot \mu \leq 0$ for all $\lambda \in \Lambda$, and so $0 \leq \mu \cdot \mu = (\sum_{\lambda \in \Lambda} \lambda) \cdot \mu \leq 0$ and thus, $\|\mu\| = 0$ implying $\mu = \vec{\mathbf{0}}$, contradicting the assumption that $\mu \neq \vec{\mathbf{0}}$.

Because $\lambda \cdot \mu > 0$ for any $\lambda \in \Lambda_\mu$, we have $\mu \in (\Lambda_\mu)^>$, and thus, Λ_μ is consistent. Then $\mu/\|\mu\|$ is in $(\Lambda_\mu)^\geq$, so the latter is non-empty.

(ii): Since $\omega^\star_{\Lambda_\mu} \in (\Lambda_\mu)^\geq \subseteq (\Lambda_\mu)^>$, we have $\omega^\star_{\Lambda_\mu} \cdot \lambda > 0$ for all $\lambda \in \Lambda_\mu$, and so, $\Lambda_\mu \subseteq \Lambda_{\omega^\star_{\Lambda_\mu}}$ (and, by definition, $\Lambda_{\omega^\star_{\Lambda_\mu}} \subseteq \Lambda$). Thus, $\Lambda_{\omega^\star_{\Lambda_\mu}}$ is non-empty. Also, $\lambda \cdot \omega^\star_{\Lambda_\mu} > 0$ for all $\lambda \in \Lambda_{\omega^\star_{\Lambda_\mu}}$, showing that $(\Lambda_{\omega^\star_{\Lambda_\mu}})^>$ is non-empty since it contains $\omega^\star_{\Lambda_\mu}$, and hence, $\Lambda_{\omega^\star_{\Lambda_\mu}}$ is consistent. □

Results in Section 7

Proposition 28. *For finite consistent set of preference inputs $\Lambda \subseteq \mathbb{R}^n$, we have the following relationships between the sets of scenarios:*

$$\omega^\star_\Lambda \in \mathrm{SI}(\Lambda) \cap \mathrm{SF}(\Lambda) \text{ and } \mathrm{SI}(\Lambda) \cup \mathrm{SF}(\Lambda) \subseteq \mathrm{SIF}(\Lambda) \subseteq \mathrm{SC}(\Lambda) = \Lambda^\geq.$$

Let $\succcurlyeq$ be any of the relations $\succcurlyeq^{mm}_\Lambda$, $\succcurlyeq^{\mathrm{C}}_\Lambda$, $\succcurlyeq^{\mathrm{I}}_\Lambda$, $\succcurlyeq^{\mathrm{F}}_\Lambda$ and $\succcurlyeq^{\mathrm{I,F}}_\Lambda$. Then, $\succcurlyeq$ is a pre-order preserved by translation and uniform positive scaling, and $\lambda \succ \vec{\mathbf{0}}$ for all $\lambda \in \Lambda$ (where $\succ$ is the strict part of $\succcurlyeq$). In addition, these relations are nested in the following ways:

$$\succcurlyeq^{\mathrm{C}}_\Lambda \subseteq \succcurlyeq^{\mathrm{I,F}}_\Lambda \subseteq \succcurlyeq^{\mathrm{I}\wedge\mathrm{F}}_\Lambda = \succcurlyeq^{\mathrm{I}}_\Lambda \cap \succcurlyeq^{\mathrm{F}}_\Lambda, \text{ and } \succcurlyeq^{\mathrm{I}}_\Lambda \cup \succcurlyeq^{\mathrm{F}}_\Lambda \subseteq \succcurlyeq^{mm}_\Lambda.$$

Proof. It follows immediately that for any $u \in \mathbb{R}^n$, the relation $\geq_u$ on $\mathbb{R}^n$ is a total pre-order that is preserved by translation and uniform positive scaling. Also, if $u \in \Lambda^\geq$ then $\lambda >_u \vec{\mathbf{0}}$ for any $\lambda \in \Lambda$.

Suppose that $S \subseteq \mathbb{R}^n$ and that $\succcurlyeq_S = \bigcap_{u \in S} \geq_u$. It follows easily that $\succcurlyeq_S$ is a pre-order (i.e., is reflexive and transitive), that is preserved by translation and uniform positive scaling (since all these properties are maintained by intersection). Furthermore, if $S \subseteq \Lambda^\geq$ then $\lambda \succ_S \vec{\mathbf{0}}$ for any $\lambda \in \Lambda$. Using this notation we have that $\succcurlyeq^{\mathrm{C}}_\Lambda = \succcurlyeq_{\Lambda^\geq}$; $\succcurlyeq^{\mathrm{I}}_\Lambda = \succcurlyeq_{\mathrm{SI}(\Lambda)}$; $\succcurlyeq^{\mathrm{F}}_\Lambda = \succcurlyeq_{\mathrm{SF}(\Lambda)}$; $\succcurlyeq^{\mathrm{I}\wedge\mathrm{F}}_\Lambda = \succcurlyeq_{\mathrm{SI}(\Lambda)\cup\mathrm{SF}(\Lambda)}$; $\succcurlyeq^{\mathrm{I,F}}_\Lambda = \succcurlyeq_{\mathrm{SIF}(\Lambda)}$, and $\succcurlyeq^{mm}_\Lambda = \succcurlyeq_{\{\omega^\star_\Lambda\}}$. Therefore, we have that each of these relations is a pre-order preserved by translation and uniform positive scaling, if $\succ$ is the strict part of any of these relations then $\lambda \succ \vec{\mathbf{0}}$ for all $\lambda \in \Lambda$.

It follows immediately from Definition 11 that $\mathrm{SF}(\Lambda) \subseteq \mathrm{SIF}(\Lambda)$ (since if $\mathbf{t}_i = 1$ for all i then $\mathrm{SF}(\Lambda_{\mathbf{t}}) = \mathrm{SF}(\Lambda)$). Theorem 25 implies that $\mathrm{SIF}(\Lambda) \subseteq \Lambda^\geq$. Also, by Theorem 7 if $u \in \mathrm{SI}(\Lambda)$ then $u \in co(\Lambda) \cap \Lambda^\geq$, which implies, by Theorem 25, that $u \in \mathrm{SIF}(\Lambda)$, putting $\mu = u$. Thus, $\mathrm{SI}(\Lambda) \subseteq \mathrm{SIF}(\Lambda)$. Therefore, $\mathrm{SI}(\Lambda) \cup \mathrm{SF}(\Lambda) \subseteq \mathrm{SIF}(\Lambda) \subseteq \Lambda^\geq$.

Definition 4 implies that $\omega^\star_\Lambda \in \mathrm{SI}(\Lambda)$, and Definition 6 implies that $\omega^\star_\Lambda \in \mathrm{SF}(\Lambda)$, using τ given by $\tau(j) = 1$ for all $j = 1, \ldots, n$, since then $\omega^\star_{\Lambda \odot \tau} \odot \tau$ just equals $\omega^\star_\Lambda$. Hence, $\omega^\star_\Lambda \in \mathrm{SI}(\Lambda) \cap \mathrm{SF}(\Lambda)$

Clearly, if $S \subseteq S'$ then $\succcurlyeq_S \supseteq \succcurlyeq_{S'}$. The last part of the result then implies $\succcurlyeq^{\mathrm{C}}_\Lambda \subseteq \succcurlyeq^{\mathrm{I,F}}_\Lambda \subseteq \succcurlyeq^{\mathrm{I}}_\Lambda \cap \succcurlyeq^{\mathrm{F}}_\Lambda$, and $\succcurlyeq^{\mathrm{I}}_\Lambda \cup \succcurlyeq^{\mathrm{F}}_\Lambda \subseteq \succcurlyeq^{mm}_\Lambda$. □

Results in Section 8

Proposition 29. *For any finite non-empty set $A \subseteq \mathbb{R}^n$ of alternatives, and any finite consistent set of preference inputs $\Lambda \subseteq \mathbb{R}^n$, $PO_{\mathrm{SF}(\Lambda)}(A) \cap PO_{\mathrm{SI}(\Lambda)}(A)$ is non-empty, and $PO_{\mathrm{SF}(\Lambda)}(A) \cup PO_{\mathrm{SI}(\Lambda)}(A) = PO_{\mathrm{SI}(\Lambda)\cup\mathrm{SF}(\Lambda)}(A) \subseteq PO_{\mathrm{SIF}(\Lambda)}(A) \subseteq PO_{\mathrm{SC}(\Lambda)}(A)$.*

Proof. From Proposition 28, we have $\omega^\star_\Lambda \in \mathrm{SI}(\Lambda) \cap \mathrm{SF}(\Lambda)$ and $\mathrm{SI}(\Lambda) \cup \mathrm{SF}(\Lambda) \subseteq \mathrm{SIF}(\Lambda) \subseteq \mathrm{SC}(\Lambda) = \Lambda^\geq$. The definition of $PO_S(A)$ implies that $PO_S(A) = \bigcup_{u \in S} PO_{\{u\}}(A)$, which is always non-empty, because A is finite, and, for any subsets S and S' of $\mathbb{R}^n$, $PO_{S \cup S'}(A) = PO_S(A) \cup PO_{S'}(A)$, and if $S \subseteq S'$ then $PO_S(A) \subseteq PO_{S'}(A)$. Putting these together gives $PO_{\{\omega^\star_\Lambda\}}(A) \subseteq PO_{\mathrm{SF}(\Lambda)}(A) \cap PO_{\mathrm{SI}(\Lambda)}(A)$; and $PO_{\mathrm{SF}(\Lambda)}(A) \cup PO_{\mathrm{SI}(\Lambda)}(A)$ equals $PO_{\mathrm{SI}(\Lambda)\cup\mathrm{SF}(\Lambda)}(A)$, and also, $PO_{\mathrm{SI}(\Lambda)\cup\mathrm{SF}(\Lambda)}(A) \subseteq PO_{\mathrm{SIF}(\Lambda)}(A) \subseteq PO_{\mathrm{SC}(\Lambda)}(A)$. Since $PO_{\{\omega^\star_\Lambda\}}(A)$ is non-empty we have that $PO_{\mathrm{SF}(\Lambda)}(A) \cap PO_{\mathrm{SI}(\Lambda)}(A)$ is non-empty. □

Proposition 30. *For finite consistent set of preference inputs $\Lambda \subseteq \mathbb{R}^n$, given any subset S of $\Lambda^\geq$, and any finite set A ($\subseteq \mathbb{R}^n$) of alternatives, Algorithm 2 returns $UND_S(A)$.*

Proof. Let Ω' be the final set Ω. Now, $\alpha' \in A - \Omega'$ implies that either α' was not included in the current Ω in Stage One, because there exists ω with $\omega \succ_S \alpha'$, or it was cut in Stage Two, because the current α strictly dominates α'. In either case, there exists $\gamma \in A$ such that $\gamma \succ_S \alpha'$ which implies that $\alpha' \notin UND_S(A)$. We have shown that $A - \Omega' \subseteq A - UND_S(A)$, and thus, $UND_S(A) \subseteq \Omega'$.

Conversely, suppose that $\alpha \in A - UND_S(A)$, so there exists some $\beta \in A$ such that $\beta \succ_S \alpha$. In fact, since A is finite and $\succ_S$ is transitive, there exists $\gamma \in UND_S(A)$ such that $\gamma \succ_S \alpha$. By the first part, $\gamma \in \Omega'$. Let us write A as $\{\alpha_1, \ldots, \alpha_h\}$, where the order reflects the order in which elements of A are chosen in the first *for loop* in the algorithm. Let Ω_i be the set Ω at the

beginning of the α_i loop. For some different i and j, $\alpha_i = \alpha$ and $\alpha_j = \gamma$ and we have $\alpha_j \succ_S \alpha_i$. Since $\Omega' \ni \alpha_j$, for all $k > j$, $\Omega_k \ni \alpha_j$. If $j < i$ then $\Omega_i \ni \alpha_j$, and thus, $\alpha_i = \alpha$ will not be included in the current Ω, i.e., $\Omega_{i+1} \not\ni \alpha$, and so $\alpha \notin \Omega'$. If $i < j$ then α_i will be removed from Ω_j at line 16, and so again $\alpha \notin \Omega'$. Thus, $A - UND_S(A) \subseteq A - \Omega'$, and hence, $\Omega' \subseteq UND_S(A)$ and $\Omega' = UND_S(A)$, proving the correctness of the algorithm. □

Proposition 31. *For finite consistent set of preference inputs $\Lambda \subseteq \mathbb{R}^n$, given any subset S of $\Lambda^{\geq}$, and any finite set A ($\subseteq \mathbb{R}^n$) of alternatives, Algorithm 3 returns $PO_S(A)$.*

Proof. Let Ψ_* be the final set Ψ, and let Ω be $(\Psi_*)^{\downarrow}$, i.e., the set returned by the algorithm. If $\alpha \in A - \Omega$ then at some stage in the algorithm, $\mathcal{F}_S(\Psi^{\downarrow}, \alpha) = NULL$. But $\mathcal{F}_S(\Psi^{\downarrow}, \alpha) = NULL$ implies that $\alpha \notin PO_S(\Psi^{\downarrow})$, which then implies that $\alpha \notin PO_S(A)$, since $\Psi^{\downarrow} \subseteq A$. We have shown that $A - \Omega \subseteq A - PO_S(A)$, and thus $PO_S(A) \subseteq \Omega$.

Conversely, suppose that $\alpha \in \Omega$, so that for some scenario $u \in S$, (α, u) is in the final set Ψ_*. It can be observed that at the end of every loop in the main algorithm, if $(\alpha, u) \in \Psi$ then for all $\beta \in \Psi^{\downarrow}$, $\alpha \cdot u \geq \beta \cdot u$. This is because when (α, u) is added to Ψ (in either line 6 or line 19) we have $u = \mathcal{F}_S(\Psi^{\downarrow}, \alpha)$; and this condition is confirmed (see line 17) whenever a new element added. In particular, we therefore have that for any $\beta \in \Omega$, $\alpha \geq_u \beta$ (i.e., $\alpha \cdot u \geq \beta \cdot u$).

Let γ be any element of A maximising $\gamma \cdot u$, so that for all $\beta \in A$, $\gamma \cdot u \geq \beta \cdot u$. Thus, $\gamma \in PO_S(A)$, so, by the first part, $\gamma \in \Omega$. The fact that (α, u) is in Ψ_* implies that $\gamma \cdot u \leq \alpha \cdot u$, and thus $\gamma \cdot u = \alpha \cdot u$. This implies that $\alpha \in PO_S(A)$, showing that $\Omega \subseteq PO_S(A)$, and hence, $\Omega = PO_S(A)$, proving the correctness of the algorithm. □

References

[1] G.N. Yannakakis, M. Maragoudakis, J. Hallam, Preference learning for cognitive modeling: a case study on entertainment preferences, IEEE Trans. Syst. Man Cybern., Part A, Syst. Hum. 39 (6) (2009) 1165–1175, https://doi.org/10.1109/TSMCA.2009.2028152.
[2] A. Birlutiu, P. Groot, T. Heskes, Multi-task preference learning with an application to hearing aid personalization, Neurocomputing 73 (7) (2010) 1177–1185, https://doi.org/10.1016/j.neucom.2009.11.025.
[3] J. Fürnkranz, E. Hüllermeier, Preference Learning, Springer, 2010.
[4] C.J. Burges, A tutorial on support vector machines for pattern recognition, Data Min. Knowl. Discov. 2 (2) (1998) 121–167.
[5] V. Vapnik, Estimation of Dependences Based on Empirical Data: Empirical Inference Science, Springer, 2006.
[6] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (3) (1995) 273–297.
[7] H. Kazawa, T. Hirao, E. Maeda, Order SVM: a kernel method for order learning based on generalized order statistics, Syst. Comput. Jpn. 36 (1) (2005) 35–43, https://doi.org/10.1002/scj.10630.
[8] R. Herbrich, T. Graepel, K. Obermayer, Support vector learning for ordinal regression, in: International Conference on Artificial Neural Networks, 1999, pp. 97–102.
[9] T. Joachims, Optimizing search engines using clickthrough data, in: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2002, ACM, New York, NY, USA, 2002, pp. 133–142.
[10] A. Stolcke, S. Kajarekar, L. Ferrer, Nonparametric feature normalization for SVM-based speaker verification, in: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2008, March 30–April 4, 2008, Caesars Palace, Las Vegas, Nevada, USA, IEEE, 2008, pp. 1577–1580.
[11] A. Ben-Hur, J. Weston, A User's Guide to Support Vector Machines, Humana Press, Totowa, NJ, 2010, pp. 223–239.
[12] A.K. Jain, R.C. Dubes, Algorithms for Clustering Data, Prentice-Hall, Inc., 1988.
[13] S. Aksoy, R.M. Haralick, Feature normalization and likelihood-based similarity measures for image retrieval, Pattern Recognit. Lett. 22 (5) (2001) 563–582, https://doi.org/10.1016/S0167-8655(00)00112-4.
[14] D. Tax, R.P. Duin, Feature scaling in support vector data descriptions, in: Learning from Imbalanced Datasets, 2000, pp. 25–30.
[15] S. Greco, V. Mousseau, R. Slowinski, Multiple criteria sorting with a set of additive value functions, Eur. J. Oper. Res. 207 (3) (2010) 1455–1470, https://doi.org/10.1016/j.ejor.2010.05.021.
[16] P. Haddawy, V. Ha, A. Restificar, B. Geisler, J. Miyamoto, Preference elicitation via theory refinement, J. Mach. Learn. Res. 4 (2003) 317–337, http://dl.acm.org/citation.cfm?id=945365.945385.
[17] D. Bouyssou, Ranking methods based on valued preference relations: a characterization of the net flow method, Eur. J. Oper. Res. 60 (1) (1992) 61–67.
[18] W.W. Cohen, R.E. Schapire, Y. Singer, Learning to order things, J. Artif. Intell. Res. 10 (1) (1999) 243–270, http://dl.acm.org/citation.cfm?id=1622859.1622867.
[19] W. Trabelsi, N. Wilson, D. Bridge, F. Ricci, Preference dominance reasoning for conversational recommender systems: a comparison between a comparative preferences and a sum of weights approach, Int. J. Artif. Intell. Tools 20 (4) (2011) 591–616, https://doi.org/10.1142/S021821301100036X.
[20] P.C. Fishburn, Lexicographic orders, utilities and decision rules: a survey, Manag. Sci. 20 (11) (1974) 1442–1471.
[21] R. Kohli, K. Jedidi, Representation and inference of lexicographic preference models and their variants, Mark. Sci. 26 (3) (2007) 380–399, http://EconPapers.repec.org/RePEc:inm:ormksc:v:26:y:2007:i:3:p:380-399.
[22] E.C. Freuder, R. Heffernan, R.J. Wallace, N. Wilson, Lexicographically-ordered constraint satisfaction problems, Constraints 15 (1) (2010) 1–28.
[23] N. Wilson, A.-M. George, B. O'Sullivan, Computation and complexity of preference inference based on hierarchical models, in: Proceedings of the 24th International Conference on Artificial Intelligence, IJCAI'15, AAAI Press, 2015, pp. 3271–3277, http://dl.acm.org/citation.cfm?id=2832581.2832705.
[24] G. Loomes, R. Sugden, Regret theory: an alternative theory of rational choice under uncertainty, Econ. J. 92 (368) (1982) 805–824.
[25] C. Boutilier, R. Patrascu, P. Poupart, D. Schuurmans, Regret-based utility elicitation in constraint-based decision problems, in: Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence, vol. 9, IJCAI 2005, Edinburgh, Scotland, 2005, pp. 929–934.
[26] D. Braziunas, C. Boutilier, Minimax regret based elicitation of generalized additive utilities, in: Proceedings of the Twenty-Third Conference on Uncertainty in Artificial Intelligence, UAI 2007, 2007, pp. 25–32.
[27] R.L. Keeney, H. Raiffa, D.W. Rajala, Decisions with multiple objectives: preferences and value trade-offs, IEEE Trans. Syst. Man Cybern. Syst. 9 (7) (1979) 403, https://doi.org/10.1109/TSMC.1979.4310245.
[28] P.C. Fishburn, Utility Theory for Decision Making, Wiley, 1970.
[29] R.L. Keeney, H. Raiffa, Decisions with Multiple Objectives: Preferences and Value Trade-Offs, Cambridge University Press, 1993.
[30] J.S. Dyer, MAUT—multiattribute utility theory, in: Multiple Criteria Decision Analysis: State of the Art Surveys, Springer, 2005, pp. 265–292.
[31] F. Bacchus, A. Grove, Graphical models for preference and utility, in: Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence, UAI 1995, Morgan Kaufmann Publishers Inc., 1995, pp. 3–10.

[32] C. Boutilier, F. Bacchus, R.I. Brafman, UCP-networks: a directed graphical representation of conditional utilities, in: Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence, UAI 2001, Morgan Kaufmann Publishers Inc., 2001, pp. 56–64.
[33] C. Boutilier, R. Patrascu, P. Poupart, D. Schuurmans, Constraint-based optimization with the minimax decision criterion, in: Principles and Practice of Constraint Programming, CP 2003, Springer, 2003, pp. 168–182.
[34] D. Braziunas, C. Boutilier, Local utility elicitation in GAI models, in: Proceedings of the Twenty-First Conference on Uncertainty in Artificial Intelligence, UAI 2005, AUAI Press, 2005, pp. 42–49, http://dl.acm.org/citation.cfm?id=3020336.3020342.
[35] C. Gonzales, P. Perny, GAI networks for utility elicitation, in: Proceedings of the Ninth International Conference on Principles of Knowledge Representation and Reasoning, vol. 4, KR2004, 2004, pp. 224–234.
[36] E. Jacquet-Lagreze, J. Siskos, Assessing a set of additive utility functions for multicriteria decision-making, the UTA method, Eur. J. Oper. Res. 10 (2) (1982) 151–164, https://doi.org/10.1016/0377-2217(82)90155-2.
[37] Y. Siskos, D. Yannacopoulos, UTASTAR: an ordinal regression method for building additive value functions, Investig. Oper. 5 (1) (1985) 39–53.
[38] G. Bous, P. Fortemps, F. Glineur, M. Pirlot, ACUTA: a novel method for eliciting additive value functions on the basis of holistic preference statements, Eur. J. Oper. Res. 206 (2) (2010) 435–444, https://doi.org/10.1016/j.ejor.2010.03.009.
[39] G. Bous, M. Pirlot, Learning multicriteria utility functions with random utility models, in: Proceedings of the Third International Conference on Algorithmic Decision Theory, ADT 2013, Bruxelles, Belgium, November 12-14, Springer, Berlin, Heidelberg, 2013, pp. 101–115.
[40] M. Grabisch, E. Raufaste, An empirical study of statistical properties of the Choquet and Sugeno integrals, IEEE Trans. Fuzzy Syst. 16 (4) (2008) 839–850.
[41] M. Grabisch, C. Labreuche, A decade of application of the Choquet and Sugeno integrals in multi-criteria decision aid, Ann. Oper. Res. 175 (1) (2010) 247–286.
[42] N. Benabbou, P. Perny, P. Viappiani, Incremental elicitation of Choquet capacities for multicriteria choice, ranking and sorting problems, Artif. Intell. 246 (2017) 152–180.
[43] C. Boutilier, R.I. Brafman, C. Domshlak, H.H. Hoos, D. Poole, CP-nets: a tool for reasoning with conditional ceteris paribus preference statements, J. Artif. Intell. Res. 21 (1) (2004) 135–191, http://dl.acm.org/citation.cfm?id=1622467.1622473.
[44] J.T. Guerin, T.E. Allen, J. Goldsmith, Learning CP-Net Preferences Online from User Queries, Springer, Berlin, Heidelberg, 2013, pp. 208–220.
[45] M. Montazery, N. Wilson, Learning user preferences in matching for ridesharing, in: Proceedings of the 8th International Conference on Agents and Artificial Intelligence - Volume 2, ICAART, INSTICC SciTePress, 2016, pp. 63–73.
[46] M. Montazery, N. Wilson, Dominance and optimisation based on scale-invariant maximum margin preference learning, in: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17, 2017, pp. 1209–1215.
[47] R. Marinescu, A. Razak, N. Wilson, Multi-objective constraint optimization with tradeoffs, in: Proceedings of the 19th International Conference on Principles and Practice of Constraint Programming, CP 2013, Uppsala, Sweden, September 16–20, 2013, Springer, Berlin, Heidelberg, 2013, pp. 497–512.
[48] N. Wilson, M. Montazery, Preference inference through rescaling preference learning, in: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016, AAAI Press, 2016, pp. 2203–2209, http://dl.acm.org/citation.cfm?id=3060832.3060929.
[49] N. Wilson, M. Montazery, Preference Inference Through Rescaling Preference Learning (extended version of IJCAI 2016 paper including proofs), Available at http://ucc.insight-centre.org/nwilson/RescalingProofs.pdf, 2016.
[50] M. Montazery, N. Wilson, Rescale-invariant SVM for binary classification, in: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17, 2017, pp. 2501–2507.
[51] C. Niculescu, L.-E. Persson, Convex Functions and Their Applications: A Contemporary Approach, Springer Science & Business Media, 2006.
[52] N. Wilson, A. Razak, R. Marinescu, Computing possibly optimal solutions for multi-objective constraint optimisation with tradeoffs, in: Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, AAAI Press, 2015, pp. 815–821, http://dl.acm.org/citation.cfm?id=2832249.2832362.
[53] L. Xia, V. Conitzer, Determining possible and necessary winners under common voting rules given partial orders, in: Proceedings of the 23rd National Conference on Artificial Intelligence, AAAI 2008, AAAI Press, 2008, pp. 196–201, http://dl.acm.org/citation.cfm?id=1619995.1620028.
[54] F. Rossi, K.B. Venable, T. Walsh, A short introduction to preferences: between artificial intelligence and social choice, Synth. Lect. Artif. Intell. Mach. Learn. 5 (4) (2011) 1–102, https://doi.org/10.2200/S00372ED1V01Y201107AIM014.
[55] N. Benabbou, P. Perny, On possibly optimal tradeoffs in multicriteria spanning tree problems, in: Proceedings of the 4th International Conference on Algorithmic Decision Theory, ADT 2015, Springer-Verlag New York, Inc., New York, NY, USA, 2015, pp. 322–337.
[56] E. Abbasnejad, S. Sanner, E.V. Bonilla, P. Poupart, Learning community-based preferences via Dirichlet process mixtures of Gaussian processes, in: Proceedings of the 23rd International Joint Conference on Artificial Intelligence, IJCAI 2013, Beijing, China, August 3–9, 2013, IJCAI/AAAI, 2013, pp. 1213–1219, http://www.aaai.org/ocs/index.php/IJCAI/IJCAI13/paper/view/6766.
[57] J. Gallier, Notes on convex sets, polytopes, polyhedra, combinatorial topology, Voronoi diagrams and Delaunay triangulations, preprint, arXiv:0805.0292, 2008.
[58] R.I. Bot, S. Grad, G. Wanka, Duality in Vector Optimization, Springer, 2009.